ATLANTIC-WIDE RESEARCH PROGRAMME ON BLUEFIN TUNA (ICCAT-GBYP - 2011)

ELABORATION OF 2011 DATA FROM SST AND THE AERIAL SURVEY ON SPAWNING AGGREGATIONS

DATA RECOVERY PLAN

Final Report

15 December 2011

Ana Cañadas, Philip Hammond, Mike Lonergan & José Antonio Vázquez

Alnilam Research and Conservation Ltd

Cándamo 116, La Berzosa, 28240 Hoyo de Manzanares, Madrid, Spain

Background

The objectives of the comprehensive ICCAT Atlantic-Wide Research Programme on Bluefin Tuna (GBYP) are to improve basic data collection and our understanding of key biological and ecological processes and to develop a robust scientific management framework. An important element of this programme is to develop fisheries independent indexes of population abundance. Therefore in 2010 and 2011 aerial surveys have been conducted in the Mediterranean on the spawning grounds.

In 2010 an analysis of the aerial survey was conducted and this included a power analysis that evaluated the ability of the survey to detect population trends in the East Atlantic and Mediterranean bluefin recovery plan. This original analysis was based on data from a single year. However, inter-annual variation (e.g. due to environmental variation and changes in population distribution) in abundance levels within areas will result in uncertainty in abundance estimates to be underestimated and the power of the survey to detect recovery to be overestimated. Data have been collected in 2011 in Areas 1, 2 and 3CM (GBYP Phase 2). This report describes the analysis for those aerial surveys in 2011.

We will focus on the value of the information obtained in the last two years despite operational and political difficulties and outline what we think is realistically necessary for a long term survey programme to succeed in its objective of providing valuable input for management.

Objectives

I. update, including 2011 data, the analysis conducted in 2010, by using the same methodological approach;

II. evaluate the sources of additional variance;

III. analyse the power to detect population trends that considers additional variance; and

IV. develop recommendations on the minimum survey design(s) required for use within a scientific management framework.

Deliverables

D1) A report for the SCRS (to be delivered **on 23th September 2011** at the latest) updating the 2010 analysis, quantifying the potential importance of additional variance and outlining the analysis to be done under GBYP Phase 3.

D2) A report for the SCRS that recommends a minimum survey design (or a range of possible survey designs) that can still meet scientific management objectives, based on a preliminary analysis (to be delivered on **23th September 2011** at the latest).

D3) A short PowerPoint presentation of the main results for the SCRS (to be delivered on **23th September 2011** at the latest).

D4) A draft final report on the effect of additional variance to be appended to a more detailed report that recommends a minimum survey design that can still meet scientific management objectives (to be delivered on **7th December 2011** at the latest).

D5) The final report, which takes into account any comments made by the GBYP Steering Committee or the ICCAT Secretariat objectives (to be delivered on **15th December 2011** at the latest).

Part I: Design-based method of analysis

1. Abundance estimates

Data

Survey design

The data for analysis were collected on aerial surveys designed by the contractors as a line transect sampling survey using software DISTANCE http://www.ruwpa.st-and.ac.uk/distance/, the "industry standard" software for line and point transect distance sampling (Cañadas, Vázquez & Hammond 2011).

Surveys were designed based on the expected available aircraft time, target survey speed, and estimated time for circling over detected schools to estimate their size. Aircraft time was allocated to each sub-area in proportion to its area. Transect lines were placed in a north-south direction to be approximately perpendicular to the coast in most blocks and to give shorter transects. Surveys were designed as equal spaced parallel lines.

Survey coverage

Figure 1 shows the original designed survey transects for sub-areas 1, 2 and 3M. Figures 2, 3 and 4 show the realised transects, the sightings made on and off effort and the effort and sightings together. Figures 5-7 show the planned and realised effort and sightings for each sub-area. Coverage of all sub-areas was comprehensive.

Data provided

Draft data collection forms were proposed by Hammond, Cañadas & Vázquez (2010) and modified in 2011. They were then generated by ICCAT. The completed data forms were provided electronically to ICCAT and passed on for analysis.

Data processing

There were a number of issues with the data forms that needed to be clarified and/or resolved prior to organising the data into an appropriate form for analysis. These included minor errors/inconsistencies and missing data. Minor errors, etc due to typographical errors and missing data were checked with the survey teams, noted and corrected in most cases. Three sightings in Baleares had to be discarded due to lack of information on cluster size and weight.

The most serious problem was, like in 2010, with the data on declination angle of each detected school (the angle to a sighted school measured from the horizontal when the school was abeam of (at 90° to) the transect line). These data are necessary to calculate the perpendicular distance data that are used to estimate detection probability. Not all the time these data were collected as intended, with inclinometers. In some cases, it was necessary to make direct measurements of perpendicular distance for all BFT sightings from the difference between the GPS positions on the transect line and over the school. The accuracy of these data is unknown. But the data finally seemed to be adequate for analysis.

When these data were collated, the result of observers not being able to see underneath the aircraft became apparent. There were few sightings close to the transect line in sub-area 3M. This is not a desirable feature of the data but has been dealt with in analysis, as described below.

The data on school size were collected in a much better and consistent way than in 2010 in all surveys. These data were recorded in two ways: estimated number of animals in the school, and estimated total weight in tons of the school. Both were used as a measure of school size in analysis, performing two analysis for each sub-area to consider both measures of school size.

Sightings made while the aircraft was transiting to and from the survey area or between transects were labelled as “off effort”. They were used to estimate the detection function, but not to estimate abundance (see methods).

A combined dataset was created that was consistent across all data fields. This dataset was entered into software DISTANCE for analysis.

Data analysis

Analysis of the data followed standard line transect methodology (Buckland *et al.* 2001).

Density of schools was estimated from the number of schools sighted, the length of transect searched and the estimated *esw* (reciprocal of the probability of detecting a school within a strip defined by the data). The equation that relates density to the collected data is:

$$\hat{D} = \frac{n\,\bar{s}}{2\,esw\,L}$$

where $\hat{D}$ is density (the hat indicates an estimated quantity), n is the number of separate sightings of schools, $\bar{s}$ is mean school size (see below), L is the total length of transect searched, and *esw* is the estimated effective strip half-width. The quantity 2 *esw* L is thus the area of the strip that has been searched. The effective strip half-width is estimated from the perpendicular distance data for all the detected animals. It is effectively the width at which the number of animals detected outside the strip equals the number of animals missed inside the strip, assuming that everything is seen at a perpendicular distance of zero. To calculate the effective strip half-width, we fitted a detection function (see below and Buckland *et al.* 2001 for further details).

Abundance was estimated as:

$$\hat{N} = A\,\hat{D}$$

where A is the size of the survey area.

Because school size was measured in tonnes in one of the analysis, the final estimate of abundance is the total estimated weight of tunas in the surveyed areas in that case.

All analysis was undertaken in software DISTANCE http://www.ruwpa.st-and.ac.uk/distance/, which estimates all quantities and their uncertainties.

Fitting the detection function

Given the large amount of sightings “off effort”, a two steps process was followed: (a) a detection function was fitted to all sightings, on and off effort; and (b) an estimate of abundance was obtained using the fitted detection function but applied only to data on effort. To do this, the MRDS (Mark-recapture distance sampling) engine in DISTANCE was used with the configuration of “single observer”.

Detection functions were fitted to the perpendicular distance data to estimate the effective strip half-width, *esw*. Multi-Covariate Distance Sampling (MCDS) methods, within the MRDS engine, were used to allow detection probability to be modelled as a function of covariates additional to perpendicular distance from the transect line. These covariates were defined in the survey design phase and included sea state, air haziness, water turbidity, observers searching, cue and estimated weight of the school. Table 1 shows the covariates tested in the models.

Analysis could not be done for each sub-area independently because of insufficient sample size for Baleares and Italy. Instead, they were stratified into two sets based on differences and similarities in the data from survey aircraft /teams. One set comprised sub-areas 1 and 2 (Baleares and Italy), and the other sub-area was 3M (Malta).

It is common practice to right truncate perpendicular distance data to eliminate sightings at large distances that have no influence on the fit of the detection function close to the transect line (the quantity of interest) but may adversely affect the fit. After initial exploration of the data, 800m right truncation distance was chosen for the Malta (3M) dataset, and no right truncation was applied to the Baleares-Italy dataset.

In sub-area 3M, lack of downward visibility beneath the aircraft meant that left truncation of the data was also necessary. Left truncation eliminates the area that has not been searched from analysis. For sub-area 3M, there were very few sightings recorded in less than 100 m from the transect line. Consequently, two analyses are explored: one with no left truncation, and one with truncation at 100m to investigate the effect on results. For sub-areas 1 and 2 no left truncation was needed.

Model diagnostics and selection

The best functional form (Half Normal or Hazard Rate model) of the detection function and the covariates retained by the best fitting models were selected based on model fitting diagnostics: AIC, goodness of fit tests, Q-Q plots, and inspection of plots of fitted functions.

Q-Q plots (quantile-quantile plots) compare the distribution of two variables; if they follow the same distribution, a plot of the quantiles of the first variable against the quantiles of the second should follow a straight line. To compare the fit of a detection function model to the data, we used a Q-Q plot of the fitted cumulative distribution function (cdf) against the empirical distribution function (edf).

For goodness of fit tests, we used the Kolmogorov-Smirnov statistic (a goodness of fit test that focuses on the largest difference between the cdf and the edf), Cramer-von Mises statistics (that focus on the sum of squared differences between cdf and edf) and the Chi-square goodness of fit statistic (that compares observed with expected frequencies of observations in each selected range of perpendicular distances).

Results

Table 2 shows the area of each survey sub-area, the number and length of searched transects, the number of sightings of bluefin tuna schools and the right truncation distances used for analysis.

For each dataset, the detection functions either using school size as weight or as number of animals are identical, and the only thing changing is the final estimate provided. Therefore we refer here as “the detection function” for each dataset, even if it was performed twice for each dataset.

Sub-areas 1 and 2

The final model selected, was null model (no covariates) with a Hazard-rate key function and no adjustment terms. The Kolmogorov-Smirnov and the Cramer-von Mises tests performed well and overall there were no significant differences between the cdf and the edf. The q-q plot shows a good agreement between the cdf and the edf. Table 3 shows the main parameters for the detection function and the results of the diagnostics tests. Figure 8 shows the fitted detection function and Figure 9 shows the Q-Q plot.

Table 4 shows the estimates of density of schools, number of individuals and total weight of bluefin tuna in each sub-area.

Sub-area 3M without left truncation

The final model selected has sea state as covariate with a Hazard-rate key function and no adjustment terms. The Kolmogorov-Smirnov test and the Cramer-von Mises tests performed well, and overall there were no significant differences between the cdf and the edf. The q-q plot shows a relatively good fit between the cdf and the edf, although there is some disagreement in the first 30% of the distances, which is obvious too when looking at the fitted detection function. The much smaller frequency of observations in the first 100m from the transect line could indicate a bias in the searching protocol of this team, or lack of visibility under the aircraft. Table 3 shows the main parameters for this detection function and the results of the diagnostics tests. Figure 10 shows the fitted detection function and Figure 11 shows the Q-Q plot.

Sub-area 3M with left truncation

The final model selected has observer as covariate with a Hazard-rate key function and no adjustment terms. The Kolmogorov-Smirnov test and the Cramer-von Mises tests performed well, and overall there were no significant differences between the cdf and the edf. The q-q plot shows a much better fit between the cdf and the edf, with no patter of disagreement. Table 3 shows the main parameters for this detection function and the results of the diagnostics tests. Figure 12 shows the fitted detection function and Figure 13 shows the Q-Q plot. As this model performed better than the model without truncation, this was chosen as the final model for sub-area 3M

Table 5 shows the estimates of density of schools, number of individuals and total weight of Bluefin tuna in sub-area 3M when sing left truncation.

All sub-areas

Table 6 pulls together the results for all sub-areas and shows results for all sub-areas combined. Overall, a total of 46,234 (CV = 40%) tonnes and 561,369 (CV = 41%) individuals of bluefin tuna was estimated in the three sub-areas.

Comparison with 2010 estimates

Table 7 shows a comparison between the estimates in 2010 and 2011 in terms of encounter rates of schools and of total weight and abundance in each of the subareas and their CVs.

Discussion

Survey logistics

The survey design generally seemed to work well. Evenly spaced north-south transects seemed to work well as a design configuration. Data collection worked much better than in 2010, although some consistency in the recording of declination angle is strongly advised.

Precision of estimates

The CV of abundance is determined by the CVs of estimated density of schools and mean school sizes in each sub-area. The CV of estimated density of schools is determined by the CVs of encounter rate (number of schools seen per survey km) and effective strip half width (*esw*). All of these quantities are functions of the number of schools seen, as well as the distribution of the data.

CVs for density of schools in all models varied between 26 % for sub-area 3M and 36 - 37% for sub-areas 1 and 2. The precision of mean school size was in the same range, between 26 and 44%. CVs for estimates of total weight were high in all sub-areas: 41% for sub-area 3M, 43% for sub-area 1 and 54% for sub-area 2. Summing over all sub-areas surveyed, the CV of total abundance was 40 %.

The number of schools seen in sub-areas 1 and 2 was insufficient to estimate an independent *esw* so data from these sub-areas were pooled. This is acceptable as long as differences in conditions in each sub-area (such as sea state, air haziness, water turbidity, observers) can be investigated as a covariate in fitting the detection function. Using the same *esw* for multiple sub-areas generates correlation in the estimates which was taken into account (in software DISTANCE) in estimating the CV of total abundance.

The main way to reduce the estimated CVs in future surveys is to increase the number of sightings. This can be achieved partly by more efficient searching (for example, using aircraft with good downward visibility) and partly by increasing the amount of searching effort (transect length). Using systematically aircraft with bubble windows would increase sample size and also avoid the need to left truncate the perpendicular distance data.

Increasing searching effort will lead to a decrease in CV of abundance but it is not possible to make exact predictions about how much. CV should improve approximately as a function of the square root of sample size, as shown in Hammond, Cañadas & Vázquez (2010). As a rough idea of the effect, if total sample size were doubled from 72 sightings to 144 sightings by improving efficiency of searching beneath the aircraft and/or increasing searching effort, we might expect the CV of total abundance to decrease from 0.33 to about 0.24.

Relative estimates of abundance

Line transect sampling assumes that detection on the transect line itself is certain. On aerial surveys, in general, it is not possible to assume this because the speed of flight means that some schools available to be sampled will inevitably not be detected (so-called perception bias). In addition, tuna spend much of their time beneath the surface and unavailable to be detected (so-called availability bias). Estimates of abundance from these surveys are thus underestimates (minimum estimates) even though a detection function has been fitted to correct for animals missed within the survey strip.

The appropriateness of these estimates as indices of abundance for the future depends on a number of factors including: timing of surveys; areas surveyed; and stability of availability and perception biases. Availability and perception bias can reasonably be assumed to be stable over time but knowledge of the distribution in time and space of bluefin tuna throughout the Mediterranean Sea is incomplete. To minimise natural variation in using survey estimates as indices of abundance over time, surveys in future years should ideally occur in the same areas at the same time of year.

Comparison with 2010 estimates

The coefficients of variation have gone down considerably in all sub-areas in 2011, when the number of sightings has increased.

In sub-area 1, there was 27% more effort in 2011 than in 2010 while there was a 57% increase in number of sightings, resulting in an increase in encounter rate (27%) and density of schools (25%). However, the mean weight of the schools has decreased 30% in 2011 with respect to 2010. Therefore, it seems that in 2011 there were more groups but smaller (in terms of weight) than in 2010, resulting in a decrease of 17% (211 tn) in final total weight for this sub-area from 2010 to 2011, which, given the wide CVs, are no significantly different.

In sub-area 2, the effort was very similar in both years but with more sightings in 2011 (67% more), resulting in a larger encounter rate (57% larger). However, density of schools is smaller (32%) in 2011. This is due to a much larger *esw* in 2011 than in 2010, so even if encounter rate of schools is larger, it refers to a much smaller searched area, and therefore when extrapolating the density within the searched area to the whole sub-area, the overall density is much larger. Also the mean weight per school was much lower in 2011 than in 2010 (66% lower). All this yields a considerably smaller total weight of Bluefin tunas in sub-area 2 with respect to 2010: 1176 tn less, representing a decrease of 76%.

Sub-area 3 had different size in 2011 due to some changes done to the limits of the block, resulting in an area 10,000 km^2 larger (around 10%). In 2011 much more effort was done in this area, more than double than in 2010, resulting also in a much larger number of observations, but in a proportional way to the increase in effort. Therefore, the encounter rates of schools remain very similar in both years. However, the *esw* in 2011 is considerably smaller than in 2010: 330m *vs* 4,830m (right truncation distance in 2011 was 800m, while it was 7,500m in 2010). This very large difference in *esw* could be explained, at least partially, with potential differences in searching protocol and/or experience, which are unknown, and should be investigated and more homogeneity would be advisable for future surveys. In addition, the mean weight per school has increased to around double in 2011 with respect to 2010, as opposed to what happened in sub-areas 1 and 2. As a consequence of all this, the total weight estimated for this sub-area is extremely larger in 2011 than in 2010 (1,820% increase). The reasons for this are unknown to us.

2. Sources of additional variance

Additional variance is the name given to the uncertainty introduced into abundance estimates by changes in the spatial distribution of animals over time. The sample variances estimated for individual sub-areas do not take into account this variability in true abundance. There is no problem if all sub-areas are surveyed in a sufficiently short period that the surveys can be considered synoptic. But if not all sub-areas are surveyed every year, the precision of estimates of total abundance summed across sub-areas surveyed in different years needs to incorporate variability in true abundance in each sub-area. If additional variance is not included in these situations, the uncertainty in estimates of total abundance will be underestimated.

Surveys for bluefin tuna in the Mediterranean Sea have thus far covered six sub-areas in 2010 and three of these same sub-areas (one of which was slightly modified) in 2011. One long-term aim could be to survey the entire Mediterranean (or the relevant part of it) to obtain the best estimates of abundance possible in any given year. In this case, there would be no need to consider additional variance. However, even if this were the aim, political and operational issues would inevitably mean that not all sub-areas would be surveyed in any given year. Realistically, it will be possible to survey only a subset of sub-areas each year. In this situation, it is necessary to consider additional variance.

Given this, if surveys for bluefin tuna in the Mediterranean Sea are to provide robust information for conservation and management it is important to attempt to quantify additional variance so that recommendations for future survey design can be as realistic as possible.

If uncertainty in abundance estimates is underestimated because additional variance is not taken into account, this will also affect the use of a time series of abundance estimates to estimate trends. Power analyses to calculate the number of surveys or the amount of survey effort required to have sufficient power to detect a given decline/recovery in abundance will generate over-optimistic results.

Work in the Scientific Committee of the International Whaling Commission has considered the issue of additional variance for many years because surveys of whale abundance are also typically composites of estimates for a number of sub-areas surveyed in different years. The method to estimate additional variance used is to fit a mixed model (using REML estimation) to the estimates of abundance in which there are fixed effects for sub-areas and random effects representing inter-annual variation in true abundance in each sub-area. The method estimates the partition of the variation in the component estimates of abundance between inter-annual trends in abundance and additional variance. This method has been used to estimate additional variance for Antarctic, North Atlantic and North Pacific minke whales (Kitakado et al. 2005; Kitakado & Okamura 2009; Bøthun et al. 2009; IWC 2010).

This approach makes the plausible assumption that the variability in the proportion of animals in each sub-area follows a Normal distribution. One possible limitation of the mixed-model approach is that it does not directly represent the real constraint that each animal must occur in one area. This should not affect the point estimates of abundance, but may affect the partitioning of uncertainty between components of the system.

The method adopted by the IWC would, in principle, also be suitable for estimation of additional variance for the GBYP surveys but there are currently only two years of survey and only three sub-areas surveyed in both years. For small datasets there is insufficient information to distinguish between a real trend in abundance and the redistribution of animals as an explanation of two estimates of abundance in a small area having non-overlapping confidence intervals. With only two estimates of each of three areas the precision of the estimates of the partitioning of the variation between inter-annual trends in abundance and additional variance will at best be very imprecise. The resulting confidence intervals can therefore also only be approximate. We are not aware of other methods that would provide estimates of additional variance for such sparse data.

Consequently, rather than attempt to estimate additional variance here, for this preliminary report we take an educated guess at what it might be and use that to inflate estimates of CVs for the purposes of running power analyses and making preliminary recommendations for future survey effort.

Estimates from the IWC work referred to above have generated estimates of additional variance that inflate the sample CVs of estimates of abundance by from 11% (for eastern North Atlantic minke whales) to 42% (for western North Pacific Bryde's whales). There is no particular reason why estimates of additional variance for bluefin tuna in the Mediterranean should be the same as for minke whales at the spatial scale of ocean basins. Nevertheless, in the absence of other information, we inflate our estimates of CV for the purposes of the power calculations and preliminary survey recommendations presented below.

For the final report under this contract, we will explore further whether there is any way to use the 2010 and 2011 data to obtain an idea of additional variance for the GBYP surveys. We will attempt to fit a minimal regression model to the data, both including an estimate of trend and under the assumption that the total population has been stable. We do not expect it to be possible to carry out any reliable model selection. We will explicitly describe the limitations and assumptions of this approach. We will also aim to make recommendations about the amount of data that would be required to be able to improve the estimates, noting that these will depend on the assumptions that are made about the plausible range of patterns that may underlie the data.

3. Power to detect population trends that considers additional variance

The power of the data to detect trends in abundance depends on a number of factors: the number of surveys; the CV of the abundance estimates; the magnitude of the trend; and the probability of a Type I error (rejecting the null hypothesis when it is actually true). Power is defined as one minus the probability of a Type II error (accepting the null hypothesis when it is actually false). The probability of a Type I error, usually referred to as α, is conventionally set at 0.05. The probability of a Type II error is usually referred to as β; hence power = 1-β. The conventional value for acceptable power is 0.8.

We used software TRENDS (Gerrodette, 1987) to investigate the relationship between power, the estimated CV of abundance, the number of surveys, and rate of population change per inter-survey period.

For sub-areas surveyed in both years, the precision of the estimate of total weight was the same in sub-area 2, which received the same amount of effort, but was lower in 2011 for sub-areas 1 and 3/3CM, reflecting the increase in survey effort (see Table 7). However, the CV of total weight summed over these three sub-areas was 0.31 in 2010 and 0.40 in 2011. The reason for the higher CV of the sum in 2011 is the very high estimate of total weight and its associated variance in sub-area 3CM.

For calculations in which we fixed the CV of abundance, we used as a baseline the CV of total abundance from the 2010 surveys (CV=0.33; Cañadas *et al*. 2010) because a larger area was surveyed in 2010 (~300,000 km2) than in 2011 (215,000 km2) and the results of the 2010 survey were more consistent (i.e. did not include the large outlier in sub-area 3CM in 2011).

We inflated the baseline CV of abundance by either 11% or 42% to represent the effect of additional variance (see above).

We also explored two scenarios for surveying larger areas in the future. The area in which it is considered useful to survey for bluefin tuna in the Mediterranean is approximately 2 million km2. We considered the CVs that would be expected in surveys of areas that were 3 times (approximating half the maximum) and 6 times (approximating the maximum) larger than the 2010 area, assuming survey coverage and bluefin tuna density were the same as in 2010. These were calculated simply by multiplying the CV by $\sqrt{3}/3$ and $\sqrt{6}/6$, respectively.

Specifically, the following power calculations were investigated:

1. The rate of recovery in abundance detectable at given power for different levels of CV of abundance as a function of the number of complete surveys. The CVs used were:

- Baseline (CV=0.33);
- Inflated to account for large additional variance (CV=0.47) and for small additional variance (CV=0.37);
- Survey area three times the 2010 area inflated for large additional variance (CV=0.27) and for small additional variance (CV=0.21);
- Survey area six times the 2010 area inflated for large additional variance (CV=0.19) and for small additional variance (CV=0.15).

2. The CV of abundance needed in future surveys to detect a trend of given magnitude as a function of the number of complete surveys. The rates of change used were 5%, 10% and 20% per survey period. Note that if it requires more than one year to complete a survey then these are rates of change per inter-survey period, not annual rates of changes.

Results are given in Tables 8 and 9.

Table 8 shows how the magnitude of the trend detectable becomes smaller as the CV of abundance decreases and the number of complete surveys increases. For the highest CVs that illustrate the baseline and its inflation for additional variance (CV=0.33-0.47), even after 10 surveys the detectable trend is only 9-12% per inter-survey period. For lower CVs representing larger survey areas (approximately half and the whole of the Mediterranean - CV=0.15-0.27), the detectable trend is 15-26% per survey period after five surveys but reduces to 4-7% per survey period after 10 surveys.

Table 9 shows how the number of surveys required to detect a given rate of recovery decreases as the CV of abundance decreases. For a relatively slow rate of recovery (5% per inter-survey period), it would take many surveys to detect this even with the most optimistic CV. For faster rates of recovery, the number of surveys required decreases. However, even for a rate of recovery of 20% per inter-survey period, it would still take 5-6 surveys with a low CV of abundance.

It is important to reiterate that these detectable trends relate to the period between complete surveys. If a complete survey takes only 1 year, the values given are annual rates of recovery. If a complete survey takes 2 years, the rates of recovery are over a 2 year period and the number of years to complete these surveys almost doubles. For example, for a CV of 0.21 (representing a survey of half the Mediterranean with small additional variance), the detectable rate of recovery is 20% with five surveys. If surveys take 2 years to complete, the detectable annual rate of recovery is (1.200.5 - 1) x 100 = 9.5% over a period of 9 years. If surveys take 3 years to complete, the detectable annual rate of recovery is (1.200.33 - 1) x 100 = 6.3% over a period of 13 years. These lower annual rates of recovery can be detected when surveys take multiple years to complete because of the greater total number of years required.

Overall, these results illustrate that if a management objective to be able to detect even a rapid rate of recovery with reasonable power, a large area (at least half the Mediterranean) needs to be surveyed for many years. In other words, a significant commitment in resources needs to be made. The amount of survey effort required to achieve this is explored below.

4. Recommendations for future survey design

The results of the power calculations above indicate the CVs of abundance that need to be achieved if particular objectives for management are to be met. These CVs can be used together with the information on actual CV from analysis of the 2010 and 2011 survey data to generate approximate levels of effort that would be needed in future surveys. In the absence of clearly defined management objectives, we present the results of some calculations for a range of scenarios.

Two particular points need to be addressed. The first is that the CV of an estimate of abundance comprises variability in a number of estimated quantities (encounter rate, average probability of detection and mean school size) and it is necessary to extract the CV of encounter rate from the overall CV. The second is that schools of tuna are not distributed randomly in space and the magnitude of encounter rate CV depends on how aggregated are the detected schools. To determine the length of transect that will generate a given CV of encounter rate therefore requires this level of aggregation to be set.

Our calculations used the following steps:

1. Use the variance of encounter rate as a proportion of variance of total abundance from the 2010 survey results to indicate what the CV of encounter rate would be for the range of CVs of total abundance explored above in the power calculations without inflation for additional variance.

2. Use the size of the survey areas used to generate the range of CVs for the power calculations, and assume that the density of bluefin tuna schools and the effective strip width searched are the same as they were on average in the 2010 survey.

3. Select the length of transect that will generate a CV of encounter rate that approximately equals the equivalent CV calculated in step 1. To do this requires making an assumption about how aggregated the sightings will be; we explored values from the 2010 survey results in which the variance of encounter rate is 3 or 5 times higher than the mean.

The value for total transect length to achieve the baseline CV of 0.33 in an area the size of that surveyed in 2010 was closer to the transect length actually flown in 2010 if the variance multiplier for encounter rate is assumed to be 5, so these results are probably more realistic. They show that to survey the whole

Mediterranean (2 million km^2) will require approximately 200,000 km of transect to achieve the selected CVs. In an area half that size, approximately 100,000 km would be required.

Table 9 shows that to detect a recovery of 10% per inter-survey period with these CVs would require 7-9 surveys depending on the size of the survey area and whether additional variance is small or large. To detect a recovery of 20% per inter-survey period with these CVs would require 5-6 surveys. If surveys took 2 years to complete, these rates of recovery would be 4.9% over 13-17 years and 9.5% over 9-11 years, respectively.

The type or form of survey design applied in 2010 and 2011 (i.e. equally spaced parallel lines) has proven to be feasible and successful and it is recommended to design future surveys in the same way.

Concluding remarks

The surveys in 2010 and 2011, although impacted by political and operational difficulties, have shown that it is feasible to collect useful data on bluefin tuna abundance in the Mediterranean Sea. Operational teething problems encountered in 2010 were addressed in a Training Workshop in February 2011. This smoothed the operation of the survey in 2011 but some aspects remain to be improved. Nevertheless, the surveys are clearly operationally feasible. Political impacts are largely beyond control and future survey design and analysis needs to take this into account. One aspect of this is the expectation that not all sub-areas will be surveyed every year, which means that additional variance will need to be taken into account. The data currently available appear insufficient to estimate the magnitude of that additional variance but further work will be done under this contract to explore this further.

The data collected have also enabled calculations to be done that give a preliminary idea of how much effort would need to be committed for a future survey programme to provide information on relative abundance of bluefin tuna that is useful for management. Useful is here taken to mean the ability to detect, with reasonable power, a recovery of a given magnitude in abundance. There is a lot of unquantifiable uncertainty but the calculations show, broadly speaking, that a survey programme of about a decade with at least 100,000km of survey effort is required to detect a recovery of around 10% per year.

References

Buckland, ST, Anderson, DR, Burnham, KP, Laake, JL, Borchers, DL & Thomas, L (2001). *Introduction to distance sampling: estimating abundance of biological populations.* Oxford University Press, Oxford.

Gerrodette, T (1987). A power analysis for detecting trends. Ecology 68: 1364-72. Software TRENDS available from http://swfsc.noaa.gov/textblock.aspx?Division=PRD&ParentMenuId=228&id=4740.

Cañadas, A, Hammond, PS & Vázquez, JA (2010). ATLANTIC-WIDE RESEARCH PROGRAMME ON BLUEFIN TUNA (GBYP - 2010) Data Recovery Plan – Elaboration of 2010 Data from the Aerial Survey on Spawning Aggregations. Final Report to ICCAT, 27 September 2010.

Cañadas, Vázquez and Hammond (2011). Atlantic-wide research programme on bluefin tuna (GBYP - 2011) - Extension of the short-term contract for the processing of aerial survey data: modification and updating of the aerial survey design for 2011 (ICCAT-GBYP phase 2). Final Report to ICCAT. March 2011.

TABLES

Table 1. Covariates tested in the models and their ranges or factor levels

Covariate	Type	Levels
Sighting related		
Cue	factor	ripples shining splash travelling other
Weight class	factor (tn)	0-10 10.1-100 100.1-200 200.1-600
School size class	factor	1-5 6-50 51-200 201-1000 1001-3000 3001-12000
Effort related		
Beaufort sea state	factor	calm (glassy) calm (rippled) smoothed (wavelets) slight
Air haziness	factor	clear slight medium heavy
Water turbidity	factor	clear slight medium heavy
Observer level	factor	9 levels
Glare intensity	factor	null slight medium strong

Table 2. Areas, number and total length of transects and number of sightings of bluefin tuna for each survey sub-area. Truncation distances are shown for each set of sub-areas.

Sub-area	Area (km^2)	Number of transects	Length of transects (km)	Number of observations (after truncation)	Left truncation (km)	Right truncation (km)
1 & 2					0	7.7
1	62,264	131	7,998	11		
2	52,461	77	8,771	10		
Subtotal 1 & 2	114,725	208	16,768	21		
3M (with left truncation)					0.1	0.8
3	100,471	65	11,429	35		
Subtotal 3M	100,471	65	11,429	35		
3M						0.8
3	100,471	65	11,429	39		
Subtotal 3M	100,471	65	11,429	39		
Total	**215,196**	**273**	**28,114**			

Table 3. Parameters and diagnostics of the detection functions.

Sub-areas	Average probability of detection (p)	Effective strip width (esw) (km)	K-S test (p)	Cramer-von Mises test (unweighted) (p)
1 and 2	0.456	3.51	0.901	0.797
3M with left truncation	0.473	0.33	0.773	0.658
3M without left truncation	0.698	0.56	0.237	0.255

Table 4. Mean school size, density and total weight and abundance of bluefin tuna in sub-areas 1 and 2.

		Sub-area		
		1	**2**	**1 and 2**
Number of transects		131	77	**208**
Transect length (km) (L)		7,997	8,771	**16,768**
Number of sightings (n)		11	10	**21**
School size (tonnes)	**Mean school size**	84.8	42.7	**64.7**
	CV (%)	26	44	**23**
School size (animals)	**Mean school size**	789	291	**552**
	CV (%)	26	31	**23**
Density of schools (per sq km)	**Density of schools**	0.00020	0.00016	**0.00018**
	CV (%)	37	36	**29**
	Lower 95% CL	0.00010	0.00008	**0.00010**
	Upper 95% CL	0.00040	0.00032	**0.00032**
Density of animals (per sq km)	**Density of animals**	0.1544	0.0472	**0.1054**
	CV (%)	43	46	**37**
	Lower 95% CL	0.0686	0.0198	**0.0522**
	Upper 95% CL	0.3477	0.1123	**0.2128**
Density of weight (per sq km)	**Density of weight**	0.0166	0.0069	**0.0122**
	CV (%)	43	54	**36**
	Lower 95% CL	0.0074	0.0025	**0.0061**
	Upper 95% CL	0.0374	0.0191	**0.0244**
Total weight (tonnes)	**Total weight**	1,033	364	**1,396**
	CV (%)	43	54	**36**
	Lower 95% CL	458	132	**696**
	Upper 95% CL	2,327	1,000	**2,801**
Total abundance (animals)	**Total abundance**	9,616	2,477	**12,092**
	CV (%)	43	46	**37**
	Lower 95% CL	4,270	1,041	**5,989**
	Upper 95% CL	21,652	5,893	**24,416**
Encounter rate of schools (per 1,000 km)	**n/L**	0.0014	0.0011	**0.0013**
	CV (%)	32	31	**23**

Table 5. Mean school size, density and total weight and abundance of bluefin tuna in sub-area 3M, using the detection function with left truncation.

		Sub-area 3M
Number of transects		65
Transect length (km) (L)		11428.9
Number of sightings (n)		35
School size (tonnes)	**Mean school size**	102.8
	CV (%)	27
School size (animals)	**Mean school size**	1,275
	CV (%)	32
Density of schools (per sq km)	**Density of schools**	0.0040
	CV (%)	26
	Lower 95% CL	0.0024
	Upper 95% CL	0.0066
Density of animals (per sq km)	**Density of animals**	5.5
	CV (%)	42
	Lower 95% CL	2.5
	Upper 95% CL	12.2
Density of weight (per sq km)	**Density of weight**	0.4463
	CV (%)	41
	Lower 95% CL	0.2024
	Upper 95% CL	0.9842
Total weight (tonnes)	**Total weight**	44,837
	CV (%)	41
	Lower 95% CL	20,331
	Upper 95% CL	98,883
Total abundance (animals)	**Total abundance**	549,276
	CV (%)	42
	Lower 95% CL	246,725
	Upper 95% CL	1,222,836
Encounter rate of schools (per 1,000 km)	**n/L**	0.0031
	CV (%)	24

Table 6. Summary of estimates for all sub-areas.

Sub-area	1	2	3M (left truncation)	Total
Survey area (km^2)	62,264	52,461	100,471	**215,196**
Number of transects	131	77	65	**273**
Transect length (km)	7,977	8,771	11,429	**28,177**
Effective strip width x 2 (km)	7.028	7.028	0.662	
Area searched (km^2)	56,062	61,643	7,566	**125,271**
% coverage	90	118	7.5	**58**
Number of schools	11	10	35	**56**
Density of schools (1000 km^{-2})	0.196	0.162	3.980	
%CV density of schools	37	36	26	
Mean weight (t)	84.8	42.7	102.8	
%CV weight	26	44	27	
Total weight (t)	**1,033**	**364**	**44,837**	**46,234**
%CV total weight	43	54	41	**40**
Mean abundance (animals)	789	291	1,275	
%CV abundance	26	31	32	
Total abundance (animals)	**9,616**	**2,477**	**549,276**	**561,369**
%CV total abundance	43	46	42	**41**

Table 7. Comparison of main results on effort, encounter rates and density of schools, and mean and total weight in the three subareas, between 2010 and 2011.

Year	**2010**			**2011**		
Sub-area	**1**	**2**	**3 (left truncation)**	**1**	**2**	**3M (left truncation)**
Survey area (km^2)	62,264	52,461	90,796	62,264	52,461	100,471
Number of transects	52	45	42	131	77	65
Transect length (km)	6,301	8,703	5,288	7,977	8,771	11,429
Effective strip width x2 (km)	9.66	2.92	9.66	7.03	7.03	0.66
Number of schools	7	6	19	11	10	35
Encounter rate of schools	0.0011	0.0007	0.0036	0.0014	0.0011	0.0031
%CV encounter rate	51	43	39	32	31	24
Density of schools (1000 km^{-2})	0.157	0.237	0.508	0.196	0.162	3.980
%CV density of schools	55	53	44	37	36	26
Mean weight (t)	127.1	124.2	50.6	84.8	42.7	102.8
%CV weight	8.0	5.6	25	26	44	27
Total weight (t)	**1,244**	**1,540**	**2,335**	**1,033**	**364**	**44,837**
%CV total weight	56	53	51	43	54	41

Table 8. The rate of recovery in abundance (trend) detectable as a function of the CV of abundance and the number of complete surveys. Probability of a Type I error (α) = 0.05. Power (1-β) = 0.80.
CV=0.33 - Baseline; CV=0.47, 0.37 - inflated to account for large and small additional variance, respectively; CV=0.27, 0.21 - survey area three times the 2010 area inflated for large and small additional variance, respectively; CV=0.19, 0.15 - survey area six times the 2010 area inflated for large and small additional variance, respectively.

CV of abundance	Number of complete surveys	Rate of recovery per survey period (%)
0.33	4	57
	5	31
	6	21
	7	16
	8	13
	9	10
	10	9
0.47	4	83
	5	44
	6	29
	7	22
	8	17
	9	14
	10	12
0.37	4	64
	5	35
	6	24
	7	18
	8	14
	9	11
	10	10
0.27	4	47
	5	26
	6	18
	7	13
	8	10
	9	9
	10	7
0.21	4	36
	5	20
	6	14
	7	11
	8	8
	9	7
	10	6

Table 8. (continuation)

CV of abundance	Number of complete surveys	Rate of recovery per survey period (%)
0.19	4	33
	5	19
	6	13
	7	10
	8	8
	9	6
	10	5
0.15	4	26
	5	15
	6	10
	7	8
	8	6
	9	5
	10	4

Table 9. Number of complete surveys needed to detect a given rate of recovery in abundance for given CV of abundance (values as in Table 10). Probability of a Type I error (α) = 0.05. Power (1-β) = 0.80.

Rate of recovery per survey period (%)	CV of abundance	Number of complete surveys
5	0.47	18
	0.37	15
	0.33	14
	0.27	13
	0.21	11
	0.19	10
	0.15	9
10	0.47	11
	0.37	10
	0.33	9
	0.27	9
	0.21	8
	0.19	7
	0.15	7
20	0.47	8
	0.37	7
	0.33	7
	0.27	6
	0.21	6
	0.19	5
	0.15	5

FIGURES

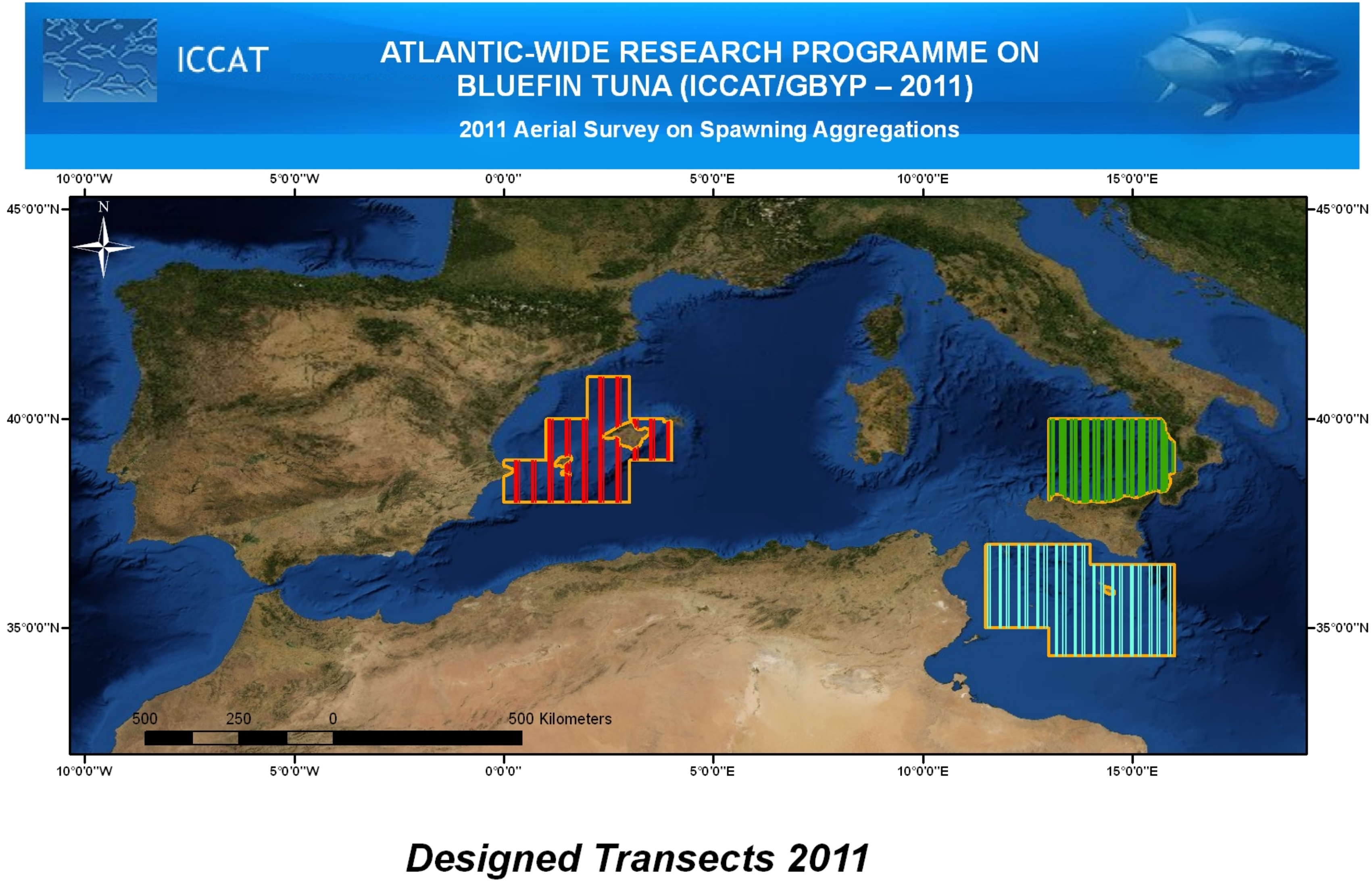


Figure 1. Originally designed transects for sub-areas 1, 2, and 3M (after Cañadas *et al.* 2011).

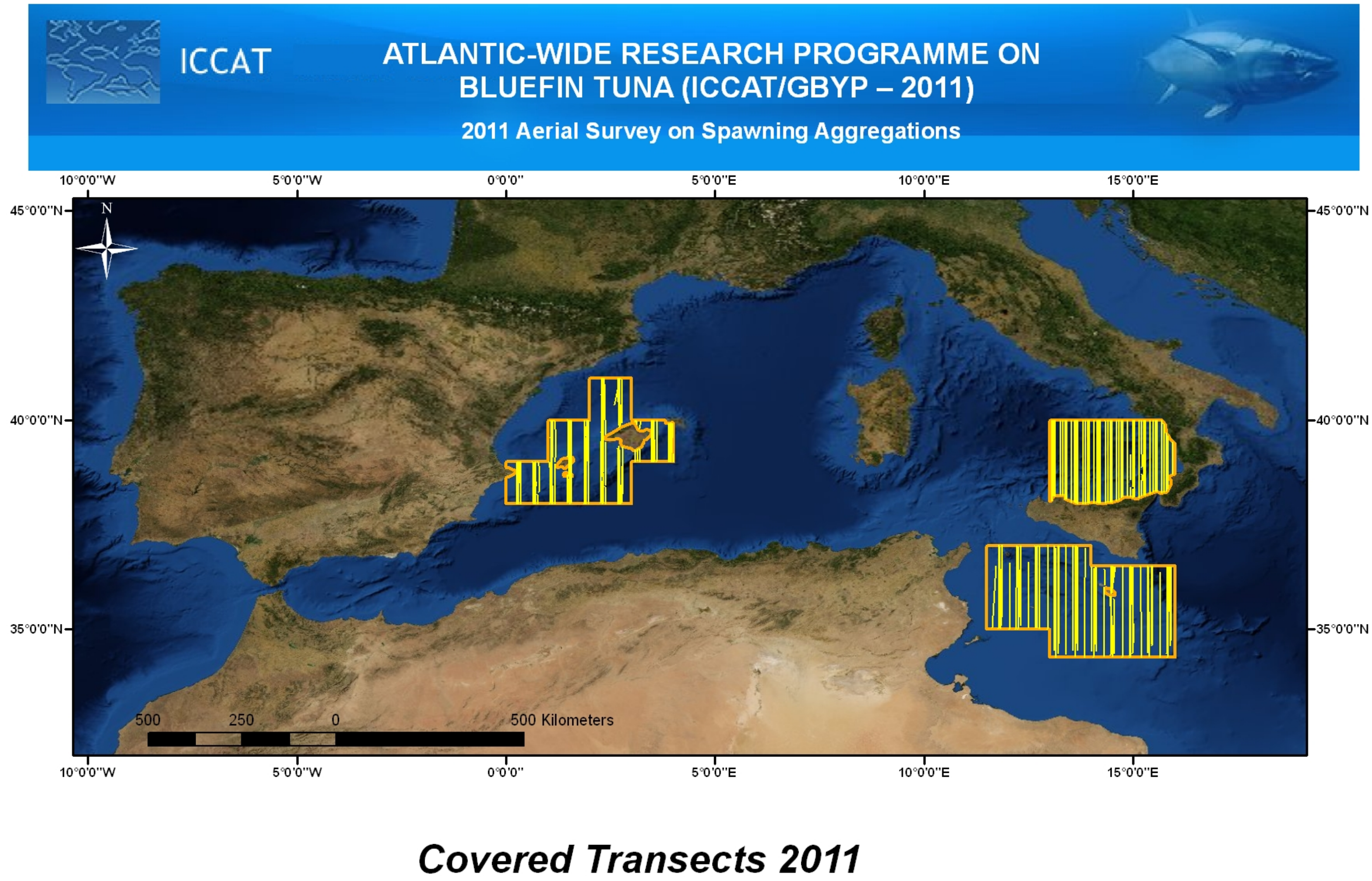


Figure 2. Transects flown on effort in sub-areas 1, 2, and 3M.

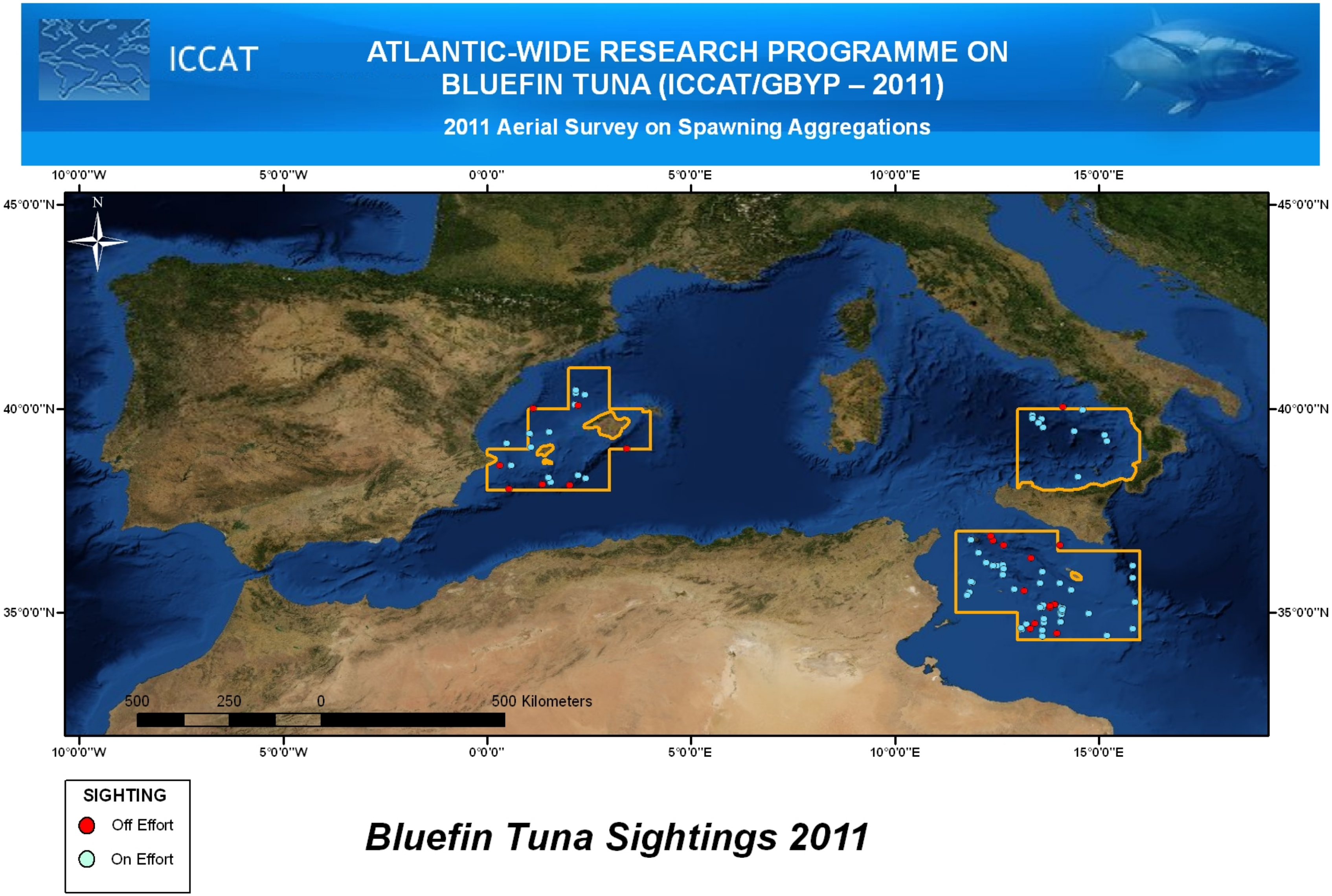


Figure 3. Sightings of bluefin tuna on and off effort in sub-areas 1, 2 and 3M.

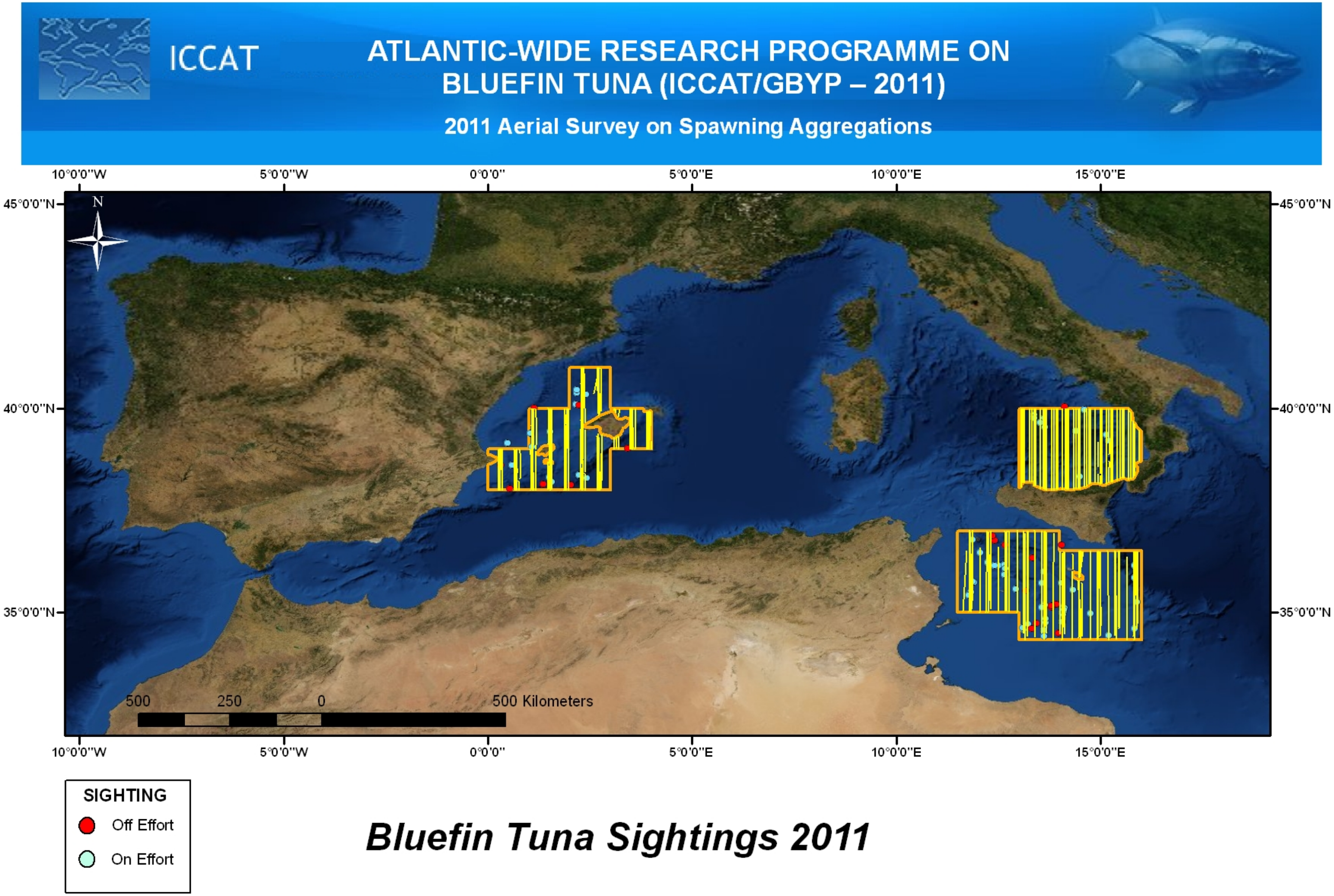


Figure 4. Transects flown and sightings of bluefin tuna on and off effort in sub-areas 1, 2 and 3M.

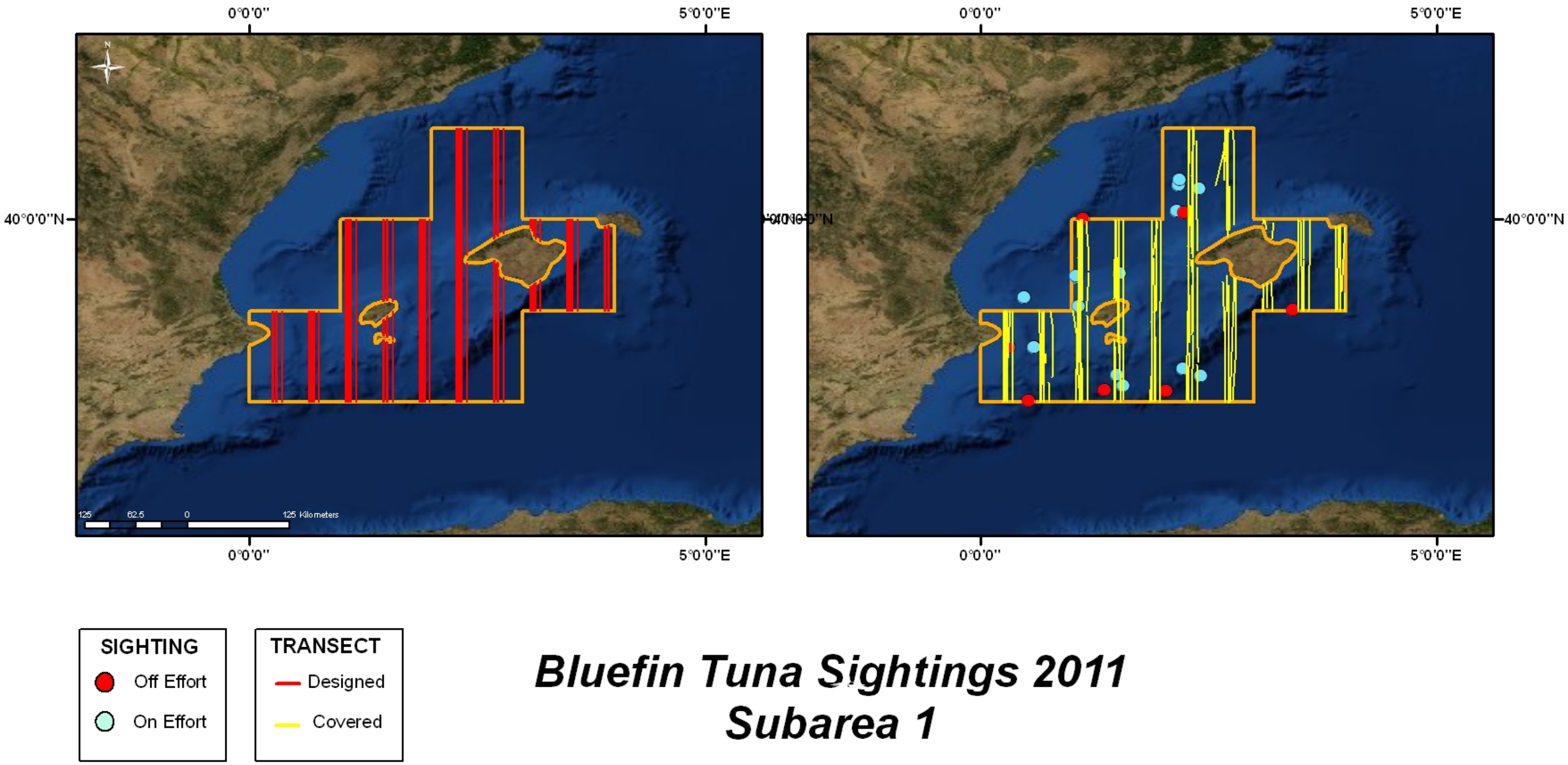


Figure 5. Transects designed and flown, and sightings of bluefin tuna on and off effort in sub-area 1.

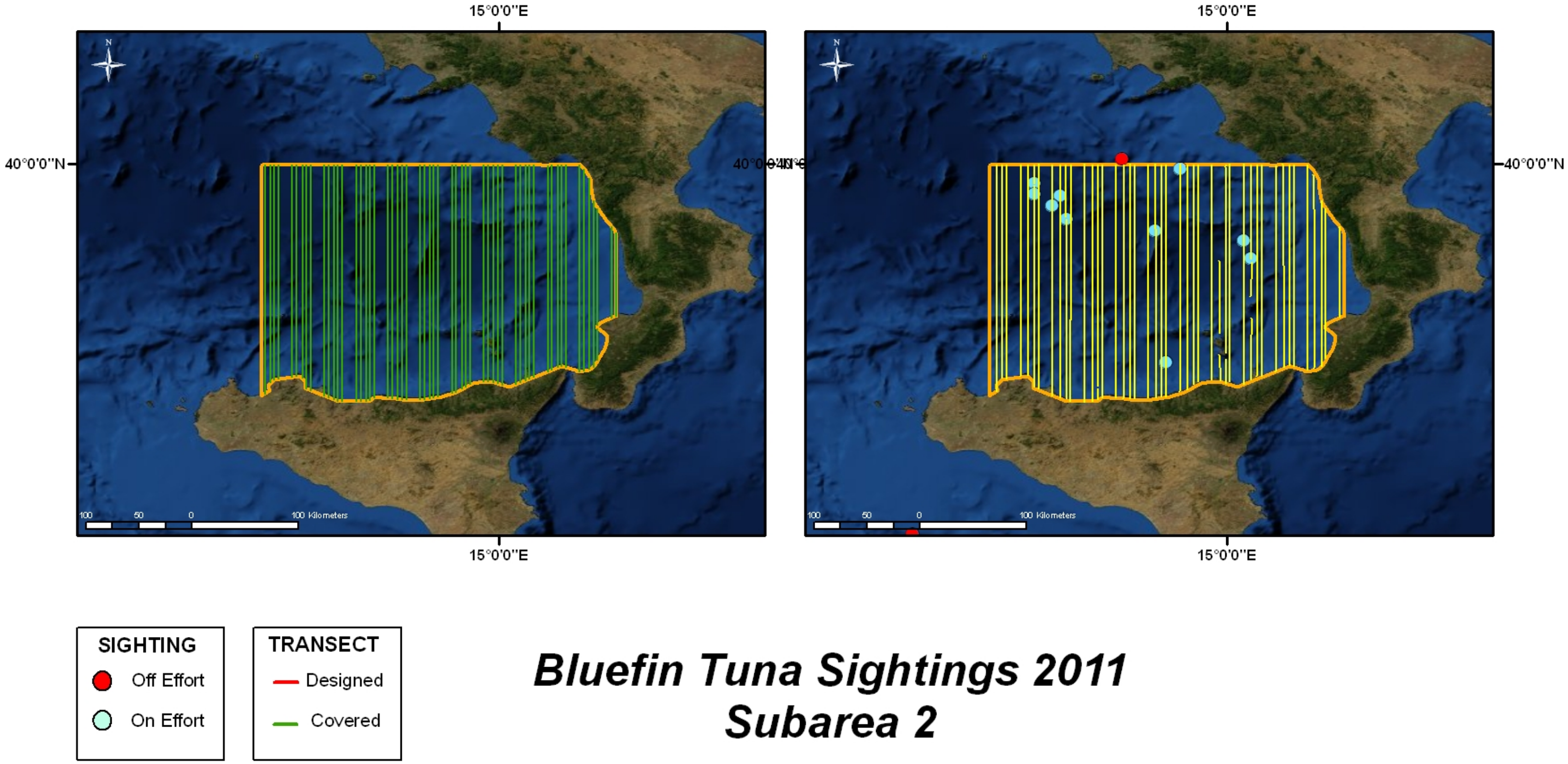


Figure 6. Transects designed and flown, and sightings of bluefin tuna on and off effort in sub-area 2.

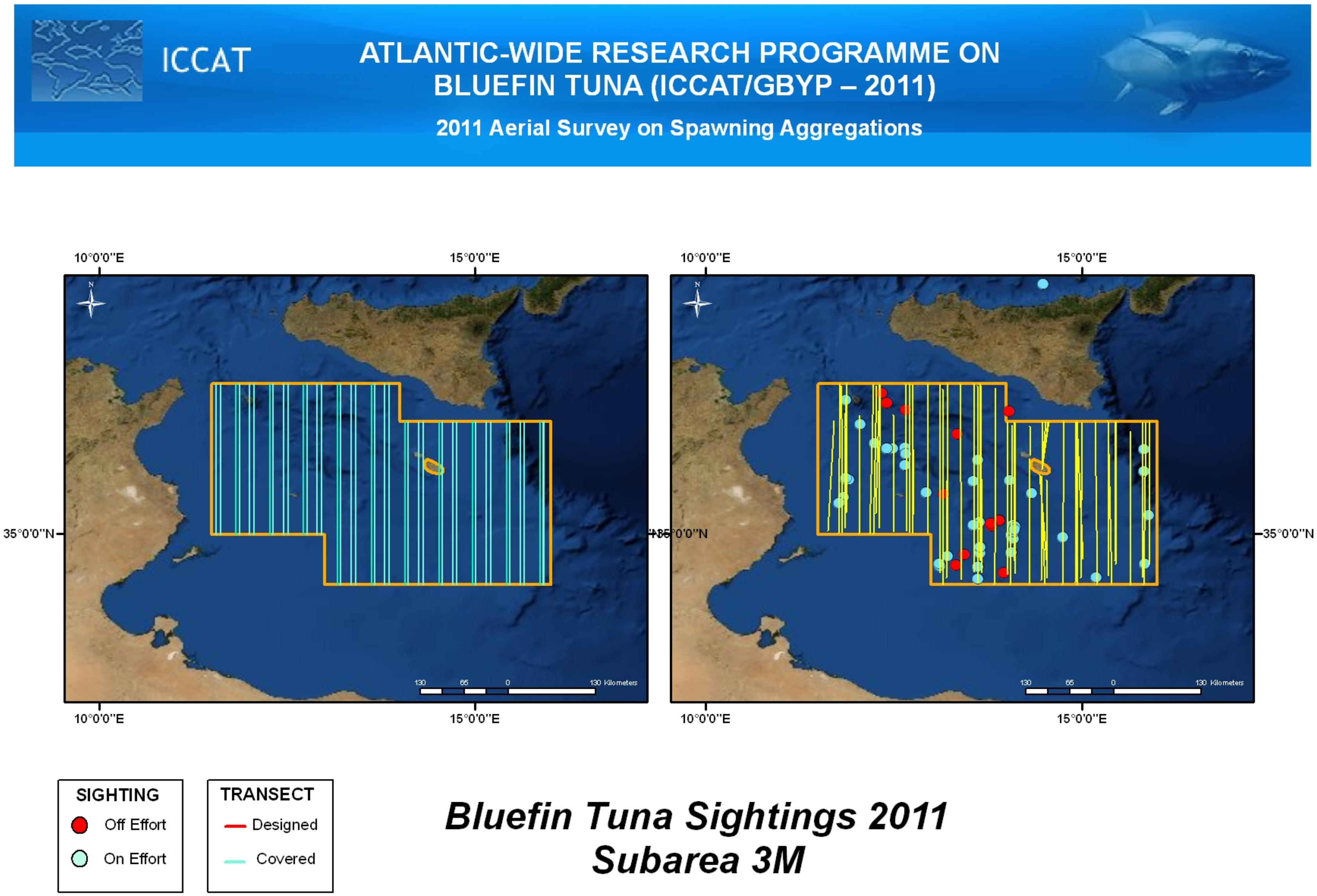


Figure 7. Transects designed and flown, and sightings of bluefin tuna on and off effort in sub-area 3M.

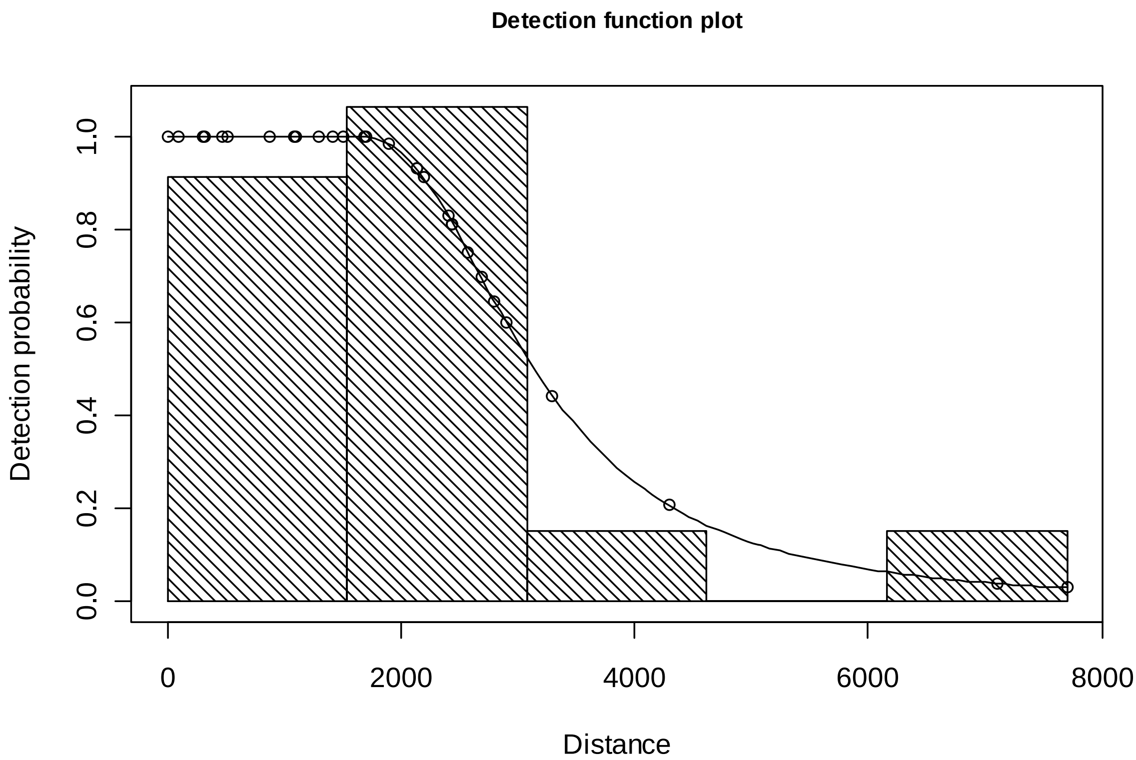


Figure 8. Detection function for sub-areas 1 and 2, scaled to 1.0 at zero perpendicular distance, and histograms of observed sightings.

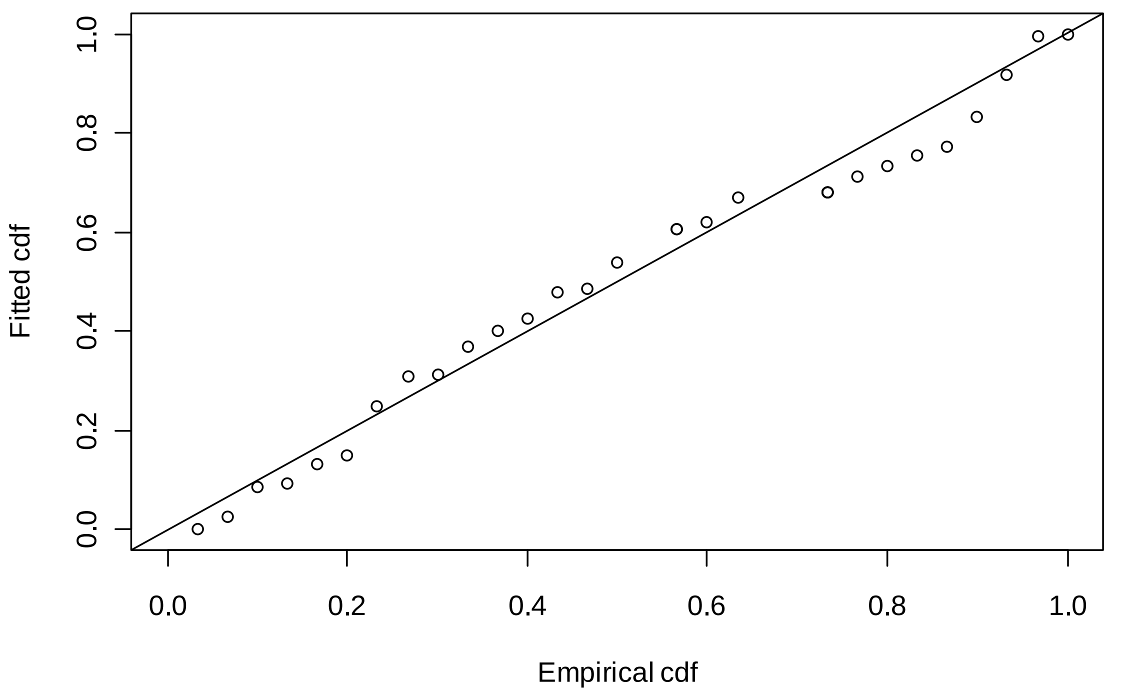


Figure 9. Q-Q plot for sub-areas 1 and 2.

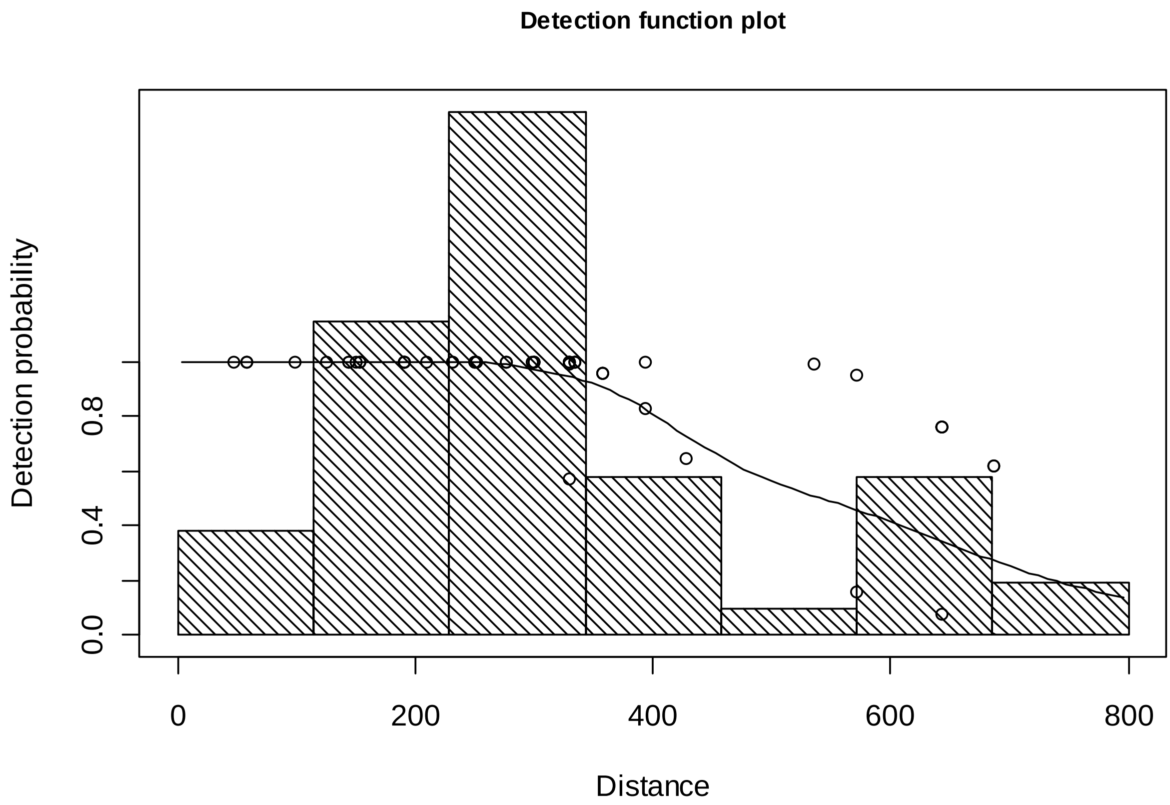


Figure 10. Detection function for sub-area 3M without left truncation, scaled to 1.0 at zero perpendicular distance, and histograms of observed sightings.

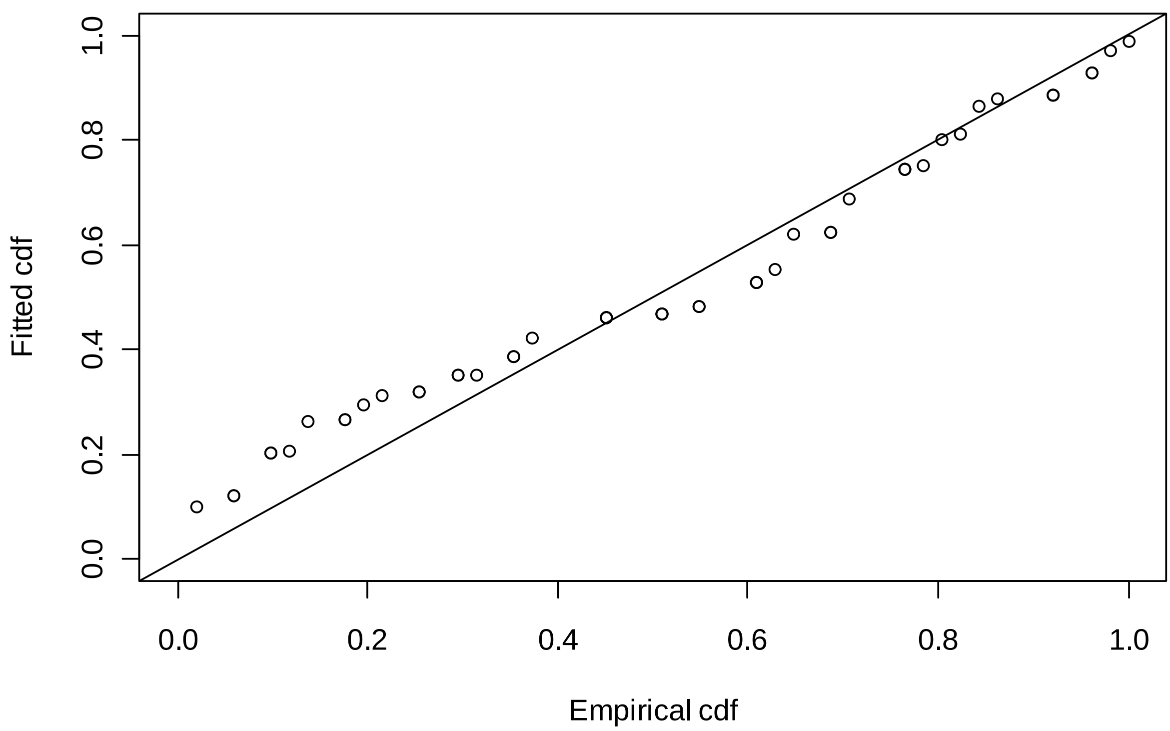


Figure 11. Q-Q plot for sub-area 3M with no left truncation

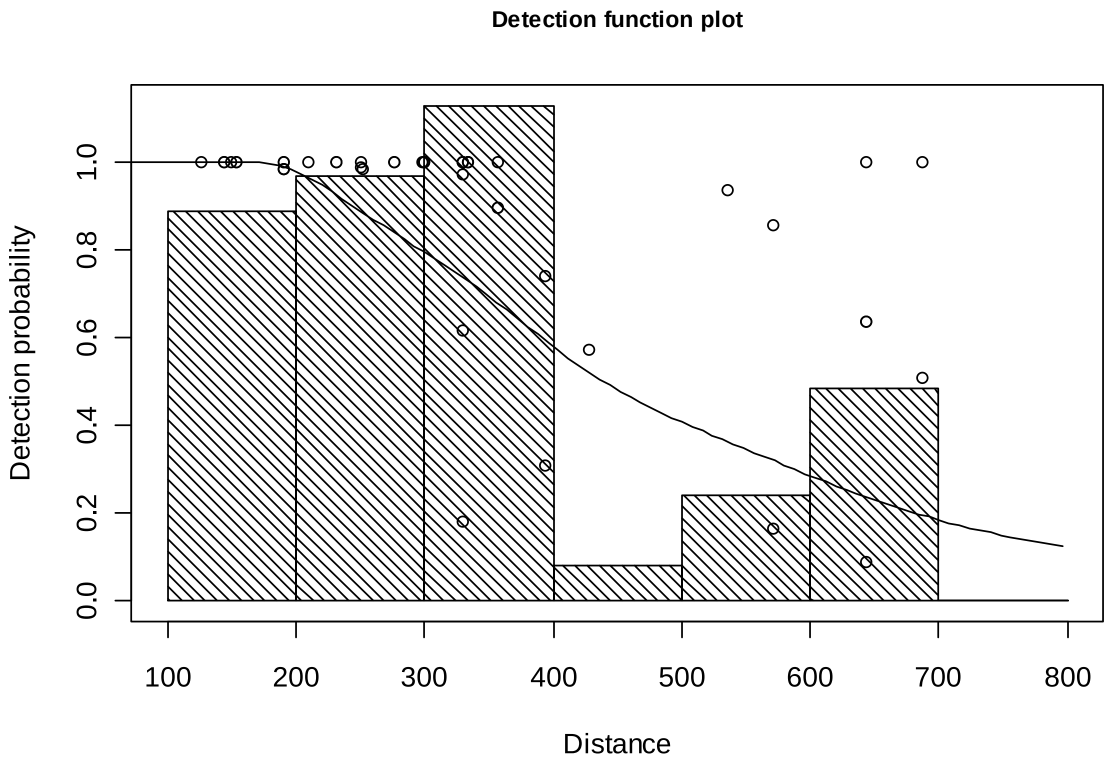


Figure 12. Detection function for sub-area 3M with left truncation at 100m, scaled to 1.0 at zero perpendicular distance, and histograms of observed sightings.

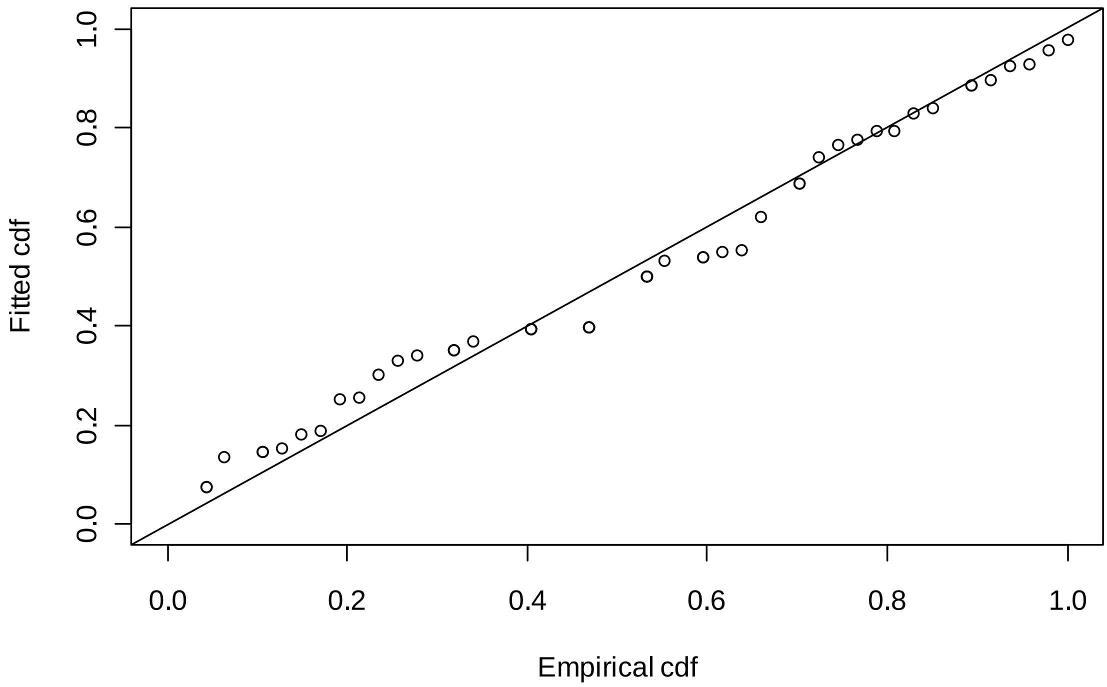


Figure 13. Q-Q plot for sub-area 3M with left truncation

Part II: Model-based method of analysis

Background

The comprehensive ICCAT Atlantic Wide Research Programme on Bluefin Tuna (GBYP) aims to improve basic data collection, understanding of key biological and ecological processes, and assessment models and management. An important element of this programme is to carry out aerial line transect surveys of the spawning population in the Mediterranean when and where schools can traditionally be sighted close to the surface to support development of fishery-independent indices.

Under the GBYP Data Recovery Framework it is desired to include an evaluation of the importance of environmental covariates, such as sea surface temperature data, in the aerial survey design. Density surface modelling is an approach that uses physical and environmental data to help explain variation in distribution and density and predict areas that are important for the focal species. When combined with line transect sampling (called the model-based method; Hedley et al. 1999), it is an alternative technique to conventional line transect sampling (design-based method; Hiby and Hammond 1989; Buckland et al. 2001).

Objectives

To fit spatial models, using methods (density surface modelling) described in Cañadas & Hammond (2006; 2008), to explore the relationship between bluefin tuna density and environmental covariates.

To provide maps of the predicted densities of bluefin tuna in the survey blocks.

Data

Data availability

The 2011 aerial survey has been made available to the authors immediately after the surveys were carried on. Sea surface temperature (sst) data were made available by ICCAT in electronic format at a resolution of 0.25°x0.25° for June, July and August 2011. Figure 1 shows the survey areas, Figure 2 the effort transects and Figure 3 shows the bluefin tuna sightings. Figures 11 to 20 (in Annex) show the mean sea surface temperature for the months of June and July and for each week in those months.

Data processing

Environmental data

A grid of cells was built during 2010 for the whole Mediterranean with the same resolution as the sst data provided (0.25ºx0.25º). These cells were populated with the sst as well as other potential covariates available to the proposers (see Table 1 for a complete list).

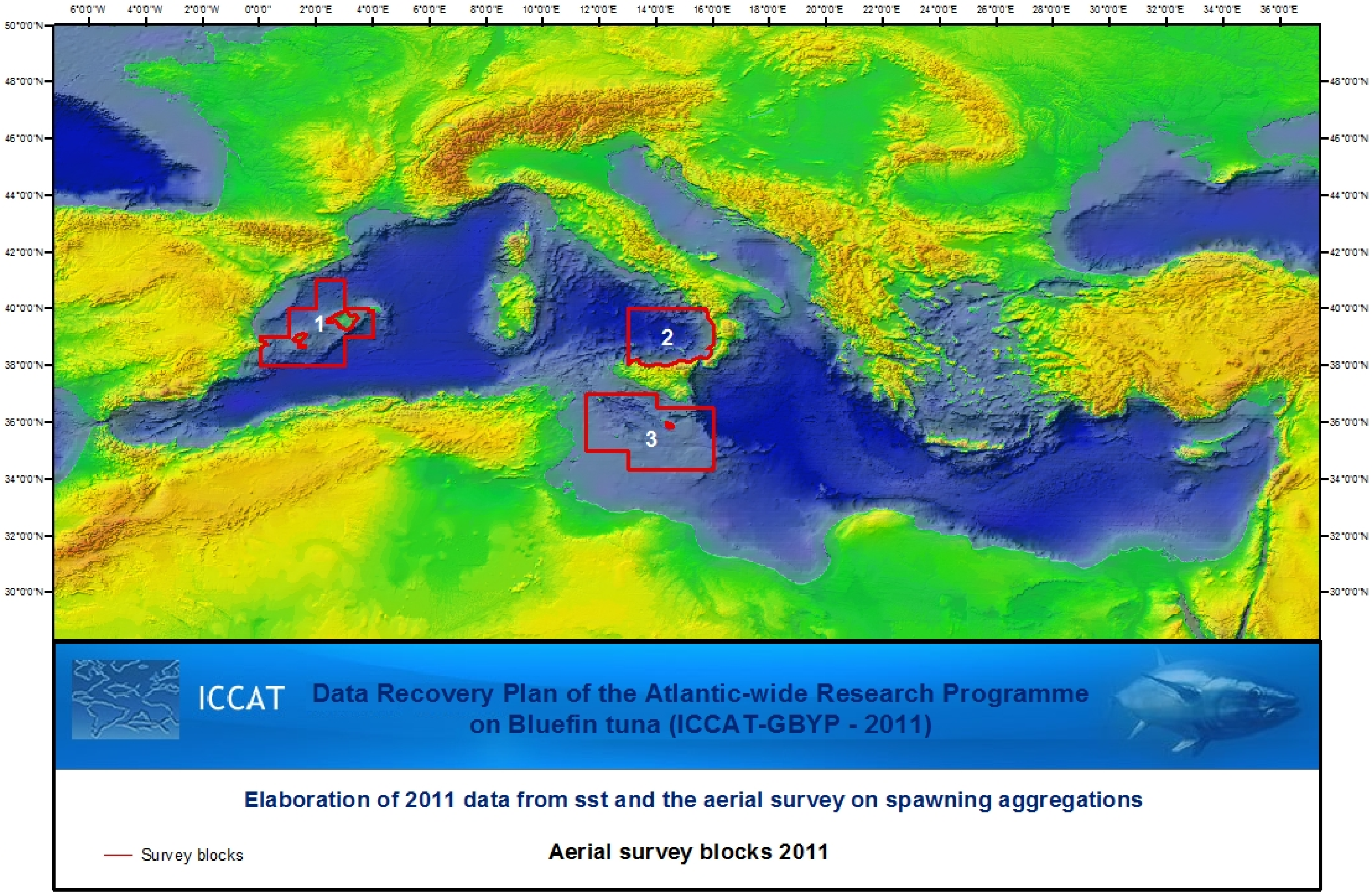


Figure 1. Aerial survey blocks considered in the analysis

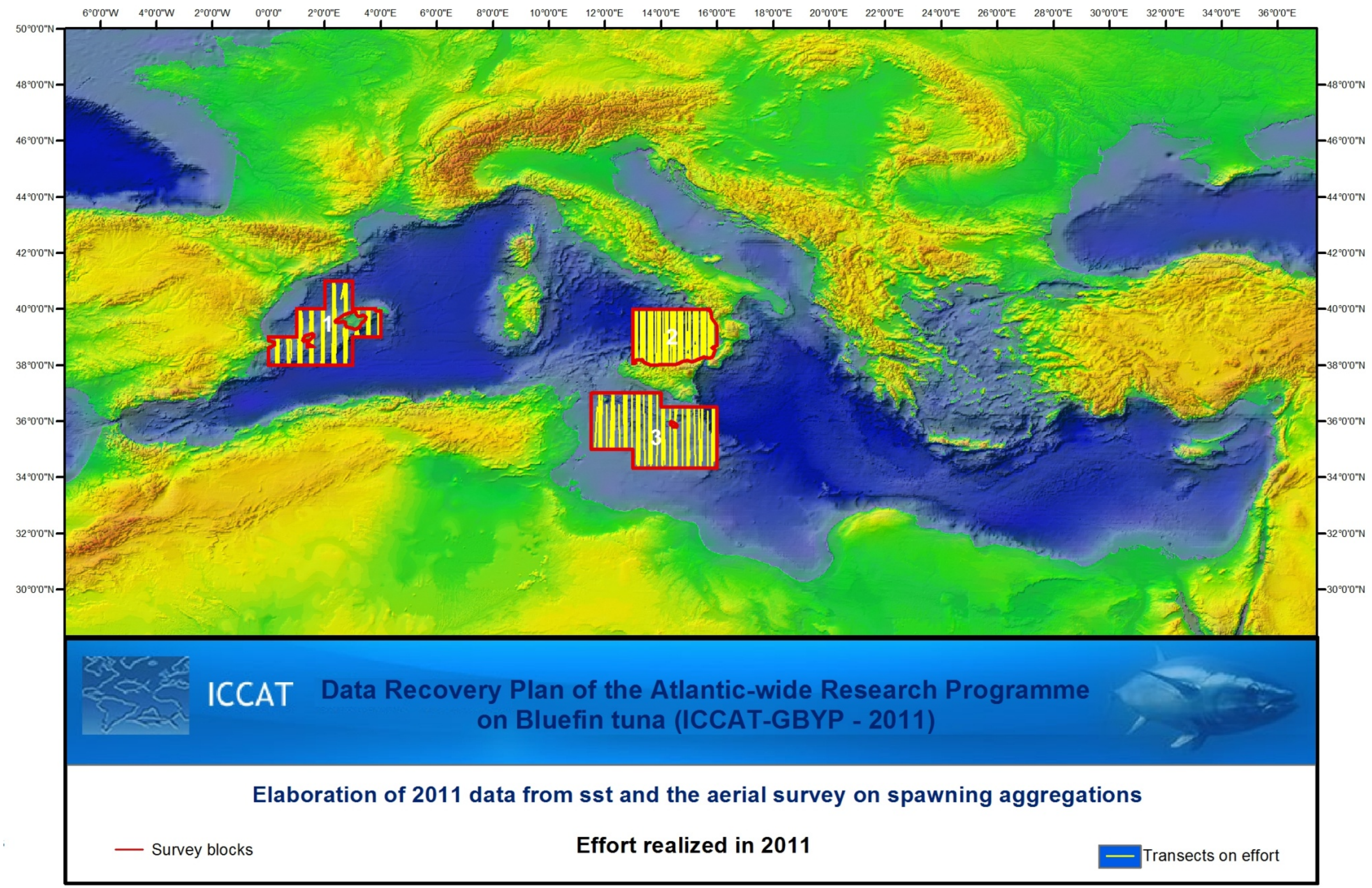


Figure 2. Aerial survey transects.

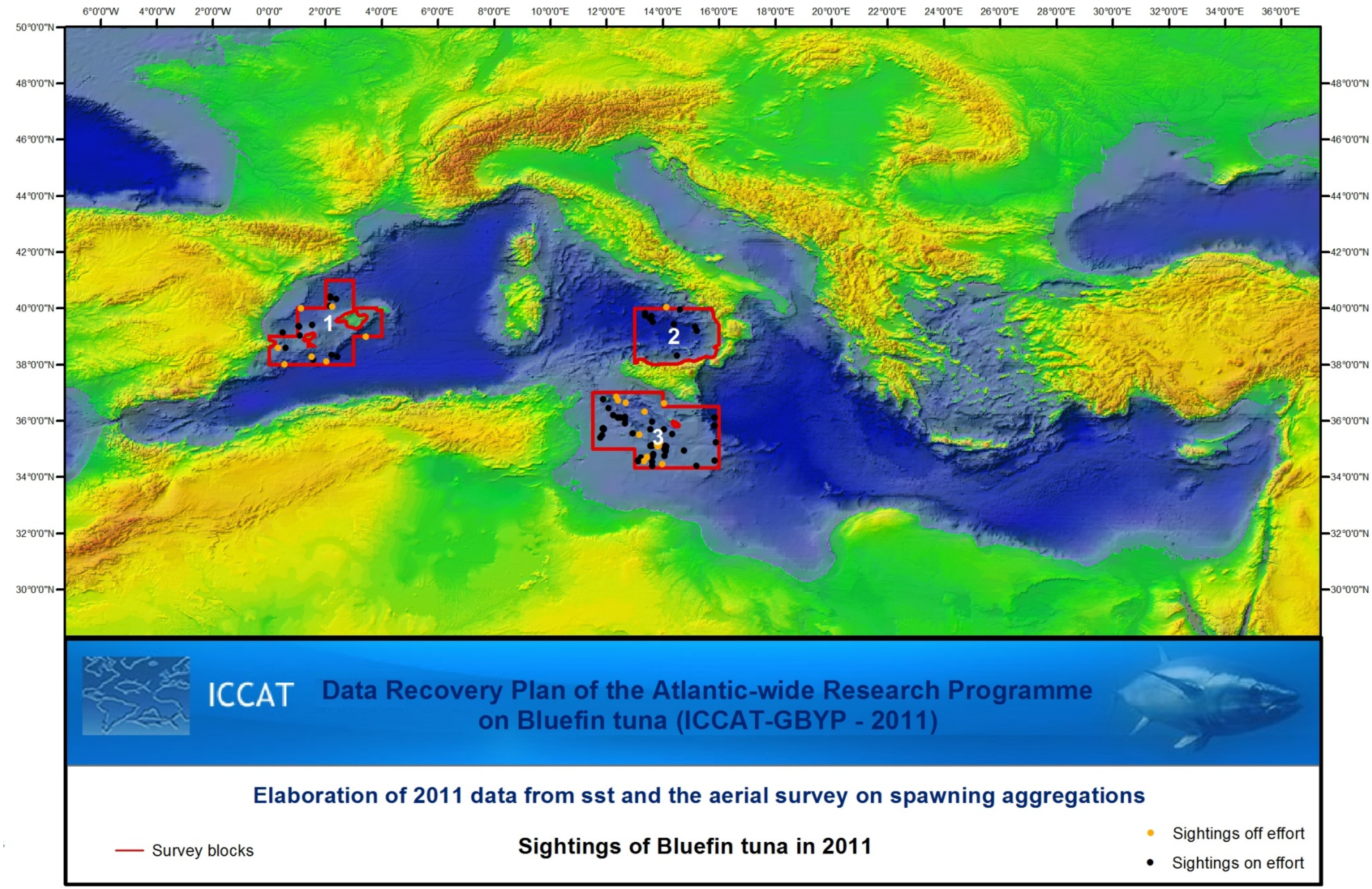


Figure 3. Sightings of bluefin tuna. Sightings in on effort are shown in black and sightings off effort are shown in orange.

Table 1. List of covariates tested in the spatial models for significance in their contribution to explain spatial distribution of sightings of bluefin tuna

Covariate	Description	Origin
Latitude	Latitude in decimal degrees	
Longitude	Longitude in decimal degrees	
Depth_mean	Mean depth within the grid cells	ETOPO 2v2 (http://www.ngdc.noaa.gov/mgg/fliers/01mgg04.html)
Depth_sd	Standard deviation of depth within the grid cells: measure of complexity of sea floor	Derived from ETOPO 2v2 data
Depth_CV	Coefficient of variation of depth within the grid cells: measure of complexity of sea floor	Derived from ETOPO 2v2 data
Ci	Contour Index: combined measure of depth and slope	(max_depth - min_depth)*100/max_depth
Dist200	Distance from the centre of the grid cell to the nearest point on the 200m depth contour	GIS
Dist1000	Distance from the centre of the grid cell to the nearest point on the 1000m depth contour	GIS
Dist2000	Distance from the centre of the grid cell to the nearest point on the 2000m depth contour	GIS
Aspect	Orientation of the sea floor relative to North	GIS
Sst_day	Sst in the grid cell on the day the segment of effort occurred	Derived from sst provided by ICCAT
Sst_week	Mean sst in the grid cell in the week the segment of effort occurred	Derived from sst provided by ICCAT
Sst_month	Mean sst in the grid cell in the month the segment of effort occurred	Derived from sst provided by ICCAT
Sst_mean3	Running average of sst on the day and the two days before the segment of effort occurred	Derived from sst provided by ICCAT
Sst_mean5	Running average of sst on the day and the four days before the segment of effort occurred	Derived from sst provided by ICCAT

Effort segments and sightings

Aerial survey effort data were organized into segments of similar length and searching conditions, which comprised the sampling units for spatial modelling. This process was done in a way so that each segment fit exactly inside one grid cell (not sharing grid cells). This process yielded a total of 1,425 segments (i.e. sampling units) of average length 21 nmi.

All segments were associated with the covariates in Table 1, according to the grid cell in which they fell.

The estimated numbers of groups (obtained through the Horvitz-Thompson estimator, see equation 1) were associated to their corresponding segment of effort (assigning 0 to the remaining segments), and this value was used as response variable for the models. Of the 1,425 segments of effort, 53 (3.7%) had associated bluefin tuna sightings, for a total of 65 sightings.

Data analysis

Spatial modelling

Generalised Additive Models (GAMs) were used to model bluefin tuna density as a function of the available covariates.

The response variable used to formulate a spatial model of abundance of groups was the estimated number of groups ($\hat{N}$) in each segment, rather than the actual counts (Hedley et al. 1999). They were estimated through the Horvitz-Thompson estimator (Horvitz & Thompson 1952), where the probability of detection was obtained from the detection function fitted to the data:

$$\hat{N}_i = \sum_{j=1}^{n_i} \frac{1}{\hat{p}_{ij}} \tag{1}$$

where n_i is the number of detected groups in the i^{th} segment, and $\hat{p}_{ij}$ is the estimated probability of the j^{th} detected group in segment i, obtained from the detection function.

The abundance of groups was modeled using a Generalized Additive Model (GAM) with a logarithmic link function. A Poisson error distribution was not considered appropriate for the response variable due to over-dispersion. Therefore, a quasi-poisson family was used, with variance proportional to the mean. The general structure of the model was:

$$\hat{N}_i = \exp\left[\ln(a_i) + \theta_0 + \sum_k f_k(z_{ik})\right] \tag{2}$$

where the offset a_i is the searched area for the i^{th} segment (calculated as the length of the segment multiplied by two times the truncation distance), θ_0 is the intercept, f_k are smoothed functions of the explanatory covariates, and z_{ik} is the value of the k^{th} explanatory covariate in the i^{th} segment.

Models were fitted using package ‘mgcv’ version 1.6-2 for R (Wood 2001). Automated model selection by a stepwise procedure was not yet implemented in the version of R used (2.11.1) (http://cran.r-project.org). Therefore, manual selection of the models was done using three indicators: (a) the GCV (General Cross Validation score) which is in practice an approximation to AIC (Wood 2000) and in which smoothing parameters (in terms of number of knots and degrees of freedom) are chosen by the software to minimize the GCV score for the model, unless they are directly specified; (b) the percentage of deviance explained; and (c) the probability that each variable is included in the model by chance. The decision to drop a term from the model was adopted following the criteria proposed by Wood (2001). In all models, a visual inspection of the residuals was also made, especially to look for trends.

The best model was used to predict bluefin tuna distribution, in a stratified fashion, within all the survey blocks.

Attempts were made also to model the weight of the schools and the school sizes as a function of the environmental covariates available, but no relationship could be found. Therefore, the estimated mean weight and the estimated mean school size of bluefin tuna per block obtained from the distance sampling analysis was used.

To obtain the final prediction of bluefin tuna weight in the survey blocks, the predicted abundance of groups in each block was multiplied by the mean weight of the block. The same was done to obtain the final prediction of bluefin tuna abundance of animals, multiplying the predicted abundance of groups in each block by the mean school size of the block.

The predictions produced by the spatial models were saved in the same grid of cells, and plotted in a G.I.S.

Results

Spatial modelling

Figures 4 to 7 show the smooth functions for the individual sea surface temperature related covariates. All means show a very similar pattern, except sst_week which shows a slightly different pattern. In all cases the trend is for higher densities towards lower temperatures, although there is a high response also around 21 to 23º for the weekly mean and around 23º for 3 and 5 day running averages. All these covariates were highly significant, but the one that better fits the data is that of sst_week (deviance explained= 10.2%, GCV=0.831), closely followed by sst_mean5 (DE=10.1%, GCV=0.833), and then sst_mean3 (DE=7.5%; GCV=0.857), and sst_day (DE=7.3%, GCV=0.859).

The best model included three covariates: latitude and longitude as an interaction and depth_mean, but none of the sea surface temperature covariates, as they did not improve the model at all. This model explained 48.1% of the deviance and these covariates were highly significant. Figure 8 shows the smooth functions for these covariates fitted in the same model.

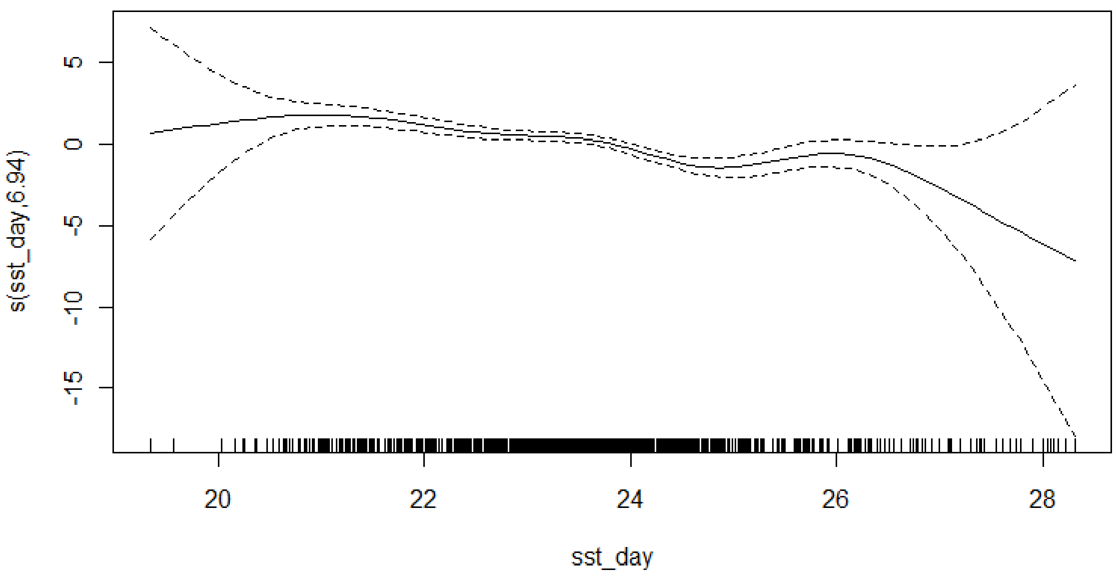


Figure 4. Smooth function for the daily sea surface temperature (sst_day). The ticks on the x axis show the distribution of the samples used in the model (the effort segments) for each covariate. The dashed lines represent ±1 se. When the line of the smooth function goes above 0 in the y axis (showing a relative index of density), it means that the covariate has a positive effect on the response variable (estimated number of groups), and *vice versa*.

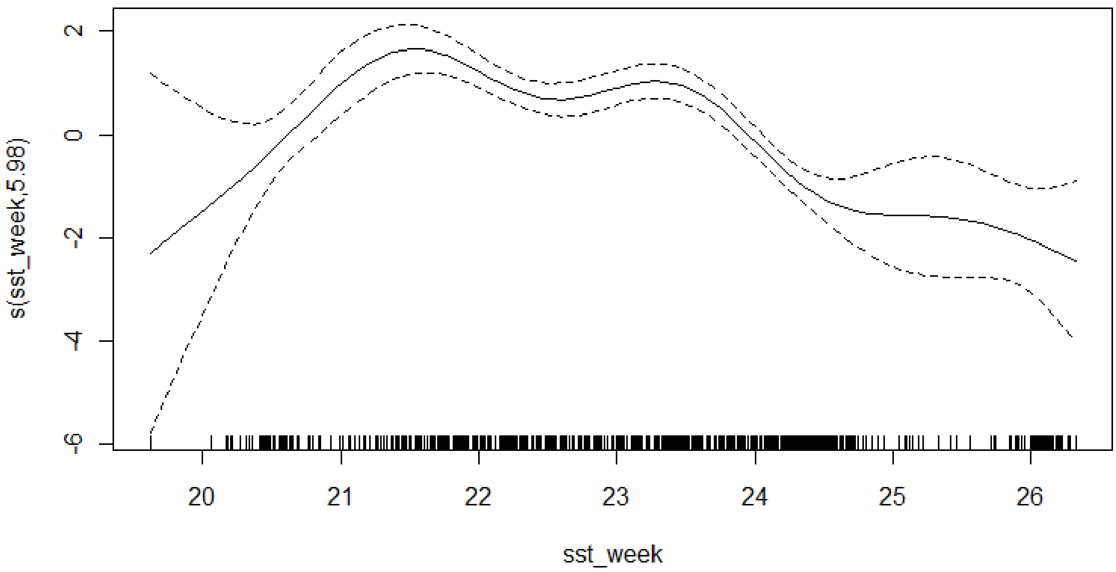


Figure 5. Smooth function for the weekly average of surface temperature (sst_week).

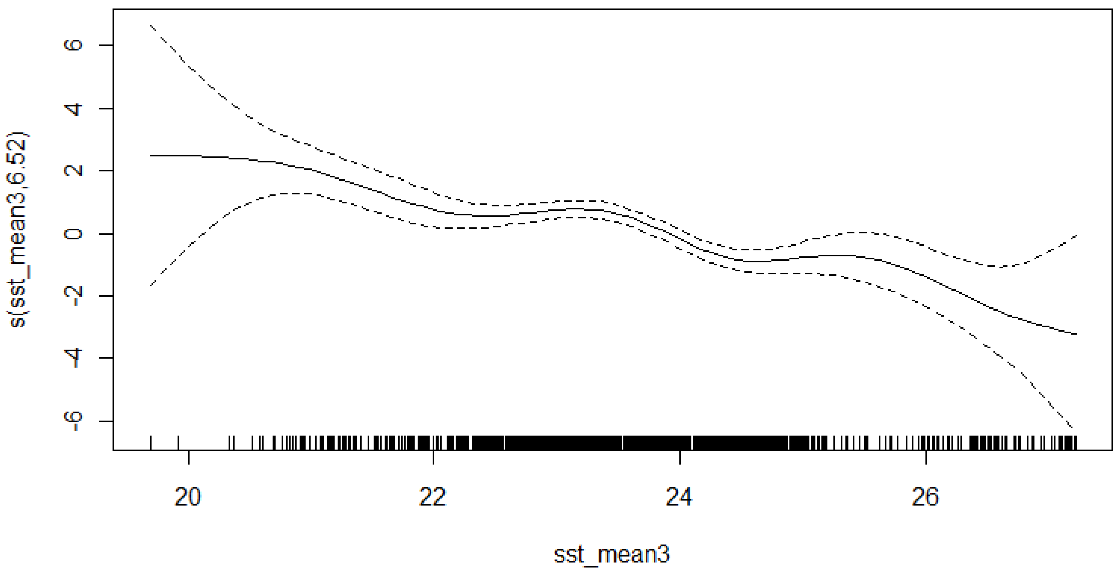


Figure 6. Smooth function for the 3-days running average of surface temperature (sst_mean3).

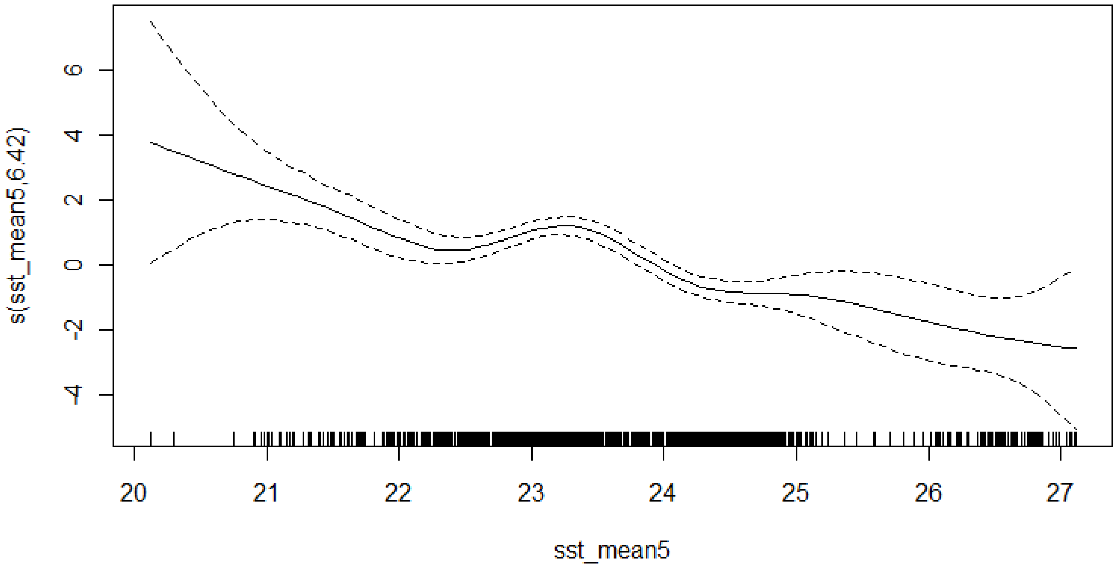
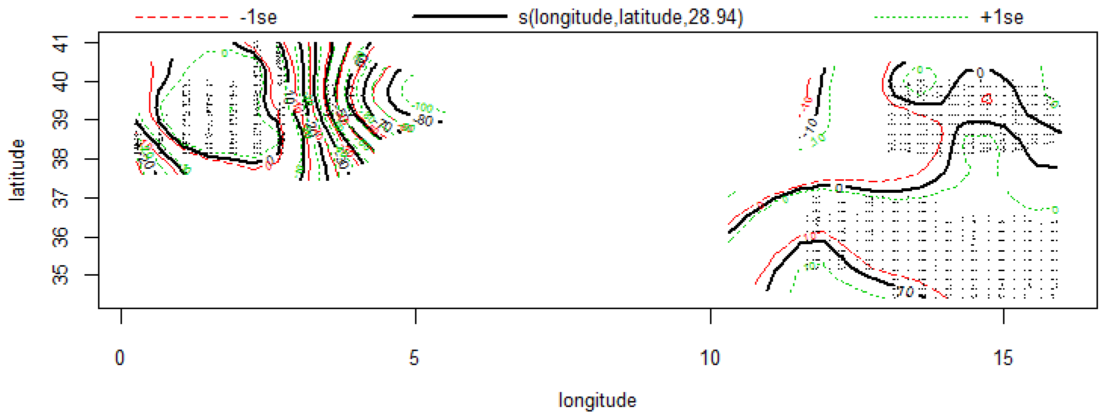


Figure 7. Smooth function for the 5-days running average of surface temperature (sst_mean5).



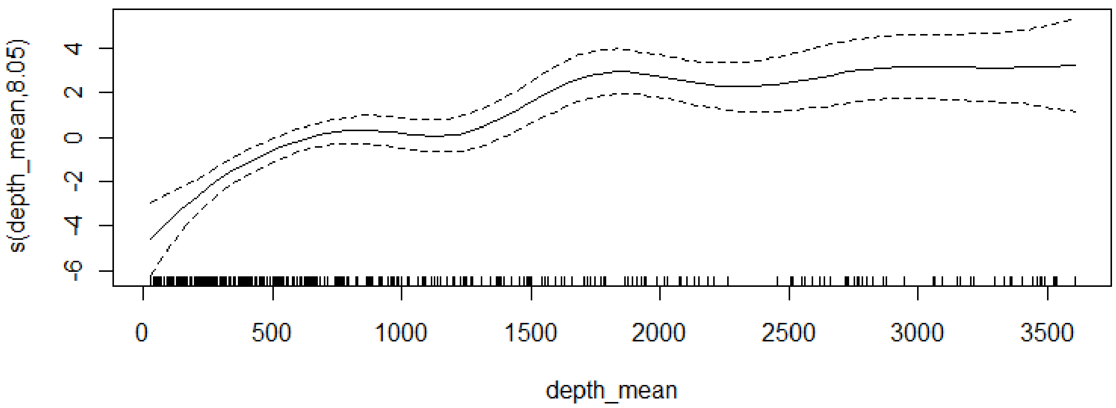


Figure 8. Smooth functions for the interaction between latitude and longitude and the depth of the sea floor (depth_mean) fitted in the same model.

Figures 9 and 10 show the prediction of density of weight and abundance of animals, respectively, from the fitted model for the three blocks of the study area.

Table 3 show the estimated total weight of bluefin tuna in each of the survey blocks predicted from the models: mean weight for the whole period, for the months of June and July, and for each of the 9 weeks of the study period. For comparison, the estimated density per block from the conventional distance sampling analysis provided in the previous contract (Hammond, Cañadas & Vázquez 2010) is also given.

Table 3. Predicted total weight (Kgs) and animal abundance of bluefin tuna in each survey block from spatial modelling (model-based method) and from conventional distance sampling (CDS, design-based method).

Block	**Mean Weight (CV)**	CDS Weight (CV)	**Mean Animal abundance (CV)**	CDS Animal Abundance (CV)
1	1,198,833 (0.583)	1,033,000 (0.429)	11,154 (0.582)	9,616 (0.429)
2	238,485 (0.679)	364,000 (0.544)	1,625 (0.605)	2,477 (0.458)
3M	51,828,826 (0.569)	44,837,000 (0.414)	642,819 (0.592)	549,276 (0.420)
Total	**53,266,144**	**46,234,000**	**655,598**	**561,369**

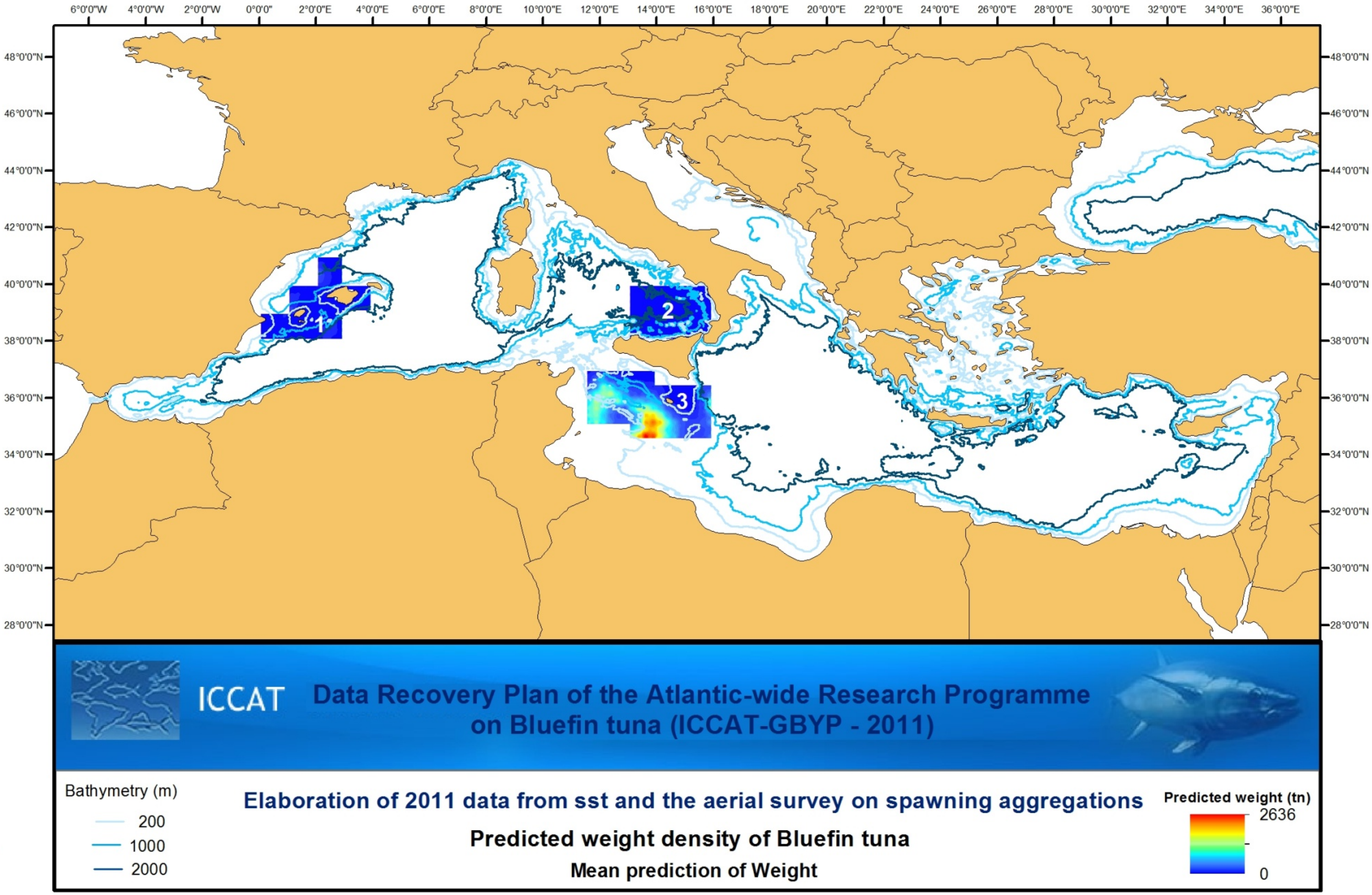


Figure 9. Predicted density of weight in tonnes of bluefin tuna in 2011

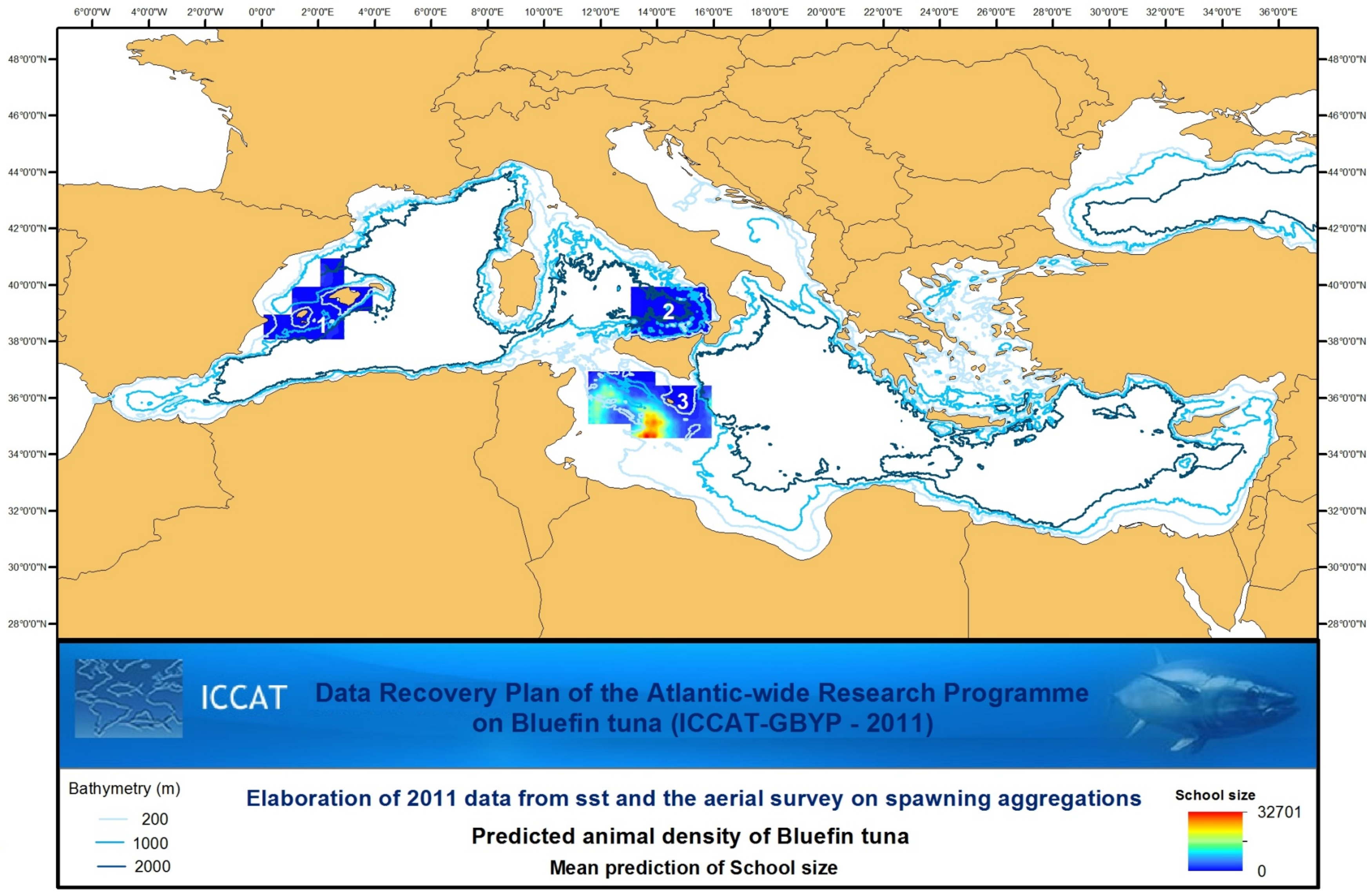


Figure 10. Predicted density of abundance of animals of bluefin tuna in 2011

Discussion on the spatial modelling

Spatial modelling to predict distribution and abundance is potentially a valuable analytical tool but its usefulness depends on the quantity and quality of the survey data. The more the survey data can be improved, the more value would be derived from the spatial modelling. Also, the greater the spatial coverage of the survey, the greater the reliability and applicability of the model results for the whole Mediterranean Sea, if this is intended. Given that there were only 3 three small areas surveyed in 2011, one should not expect spatial modelling to be able to provide very informative results in this case and.

In the analysis of 2011 data, none of the sea surface temperature covariates were selected for the final model, in contrast to the model for 2010. This may be a result of the fact that there is much less weaker contrast in sst SST over the survey areas in 2011 because, given that there are less were fewer sub-areas surveyed, and they were relatively close together. These are much closer by, yielded the non-usefulness of these covariates in 2011.

Nevertheless, when fitted as unique covariates, all of the SST covariates them were significant, but interestingly, the pattern shown in 2011 is opposite to that showed in 2010, i.e., in 2011 density decreased with sst, while it increased in 2010. The reasons for this change of pattern are unknown, but maybe it may be is also related with to the lack of survey data from in the high density easternmost (and warmer) areas of the Mediterranean, which were surveyed in 2010 and showed higher densities. Notwithstanding this, in both years there seems to be a consistent peak in predicted density around 23ºC, especially when considering the time lags of 3 and 5 days.

However, a large caveat in the analysis of the 2011 data is that there were apparently problems with the recording of the perpendicular distance in sub-area 3M (Malta); estimated densities in this area were yielding unrealistically high and the results are therefore questionable. Therefore, we do not consider appropriate to extract any conclusion from the results for this area.

If the data from sub-area 3M are reliable then the estimated weight for sub-area 1 (Baleares) is consistent with those from 2010 (Conventional Distance Sampling (CDS) 2010 = 1,244 tn; CDS 2011= 1,033tn) and there seems to be a decrease in sub-area 2 (Sicily) from 1,540 tn in 2010 to 0,364 tn in 2011. However, if the data from sub-area 3M are not reliable then no conclusions can be made for sub-areas 1 and 2 either because the fitted model will be incorrect. Omitting sub-area 3M from the spatial modeling left too small a coverage for spatial modeling to be worthwhile.

The coefficients of variation of the model-based method in 2011 are larger than for the design-based method of estimating abundance, probably due to small sample size and the lack of enough good environmental covariates to account for the variability. Hence, the results of the design-based method are more precise and therefore more useful.

If future surveys are conducted with greater and more widespread coverage, spatial modelling could provide very valuable results, as suggested by the analysis of the 2010 data.

Additional comments on additional variance

In our report of 23 September 2011, we discussed sources of additional variance and stated that we would explore further whether there is any way to use the 2010 and 2011 data to obtain an idea of additional variance for the GBYP surveys. We indicated that we would attempt to fit a minimal regression model to the data. However, after reconsideration of the survey results, it is clear that even this limited analysis is beyond the data available. We therefore describe the patterns of variability in the results for sub-areas surveyed in both 2010 and 2011, outlining why further analysis is unwarranted at this time. We also discuss the relevance of the results in the recent paper by Druon et al. (2011) to the estimation of additional variance in these data.

<u>Comparison of 2010 and 2011</u>

Comparing the pairs of estimates for each subarea (Table 7 of report of September 2011) shows very different patterns, and makes it difficult to have confidence that any pattern of interannual variability extrapolated from them can be applied across the whole population.

The total weights of bluefin tuna estimated for sub-area 1 in 2010 and 2011 are similar. The two values lie well within each other's confidence intervals, and are consistent with this sub-area containing a closed and stable population. Unless it is believed that this population tripled or halved during that year, it would be difficult to interpret these results as indicating any interannual additional variance. In principle it would be possible to put an upper bound on the possible effect of additional variance on the population, though this would depend on additional assumptions. A rough indication of the limits on such an approach here comes from the fact that there is a 5% chance that the absolute difference in two values drawn from one Normal distribution are less than 1/12 of the standard deviation apart. In this context it would suggest that the uncertainty due to additional variance in the estimates of the total weight of bluefin tuna in sub-area 1 is very unlikely to be more than 2,400 tonnes, twice the point estimates.

Sub-area 3M is at the other extreme. The 2010 survey puts the upper limit of the 95% confidence interval on the total weight of bluefin tuna in this area at around 8,000 tonnes. The equivalent lower bound on the 2011 estimate is 20,000 tonnes. If these figures are accurate reflections of the abundance at those times, then either the population has increased very rapidly or there is a very high rate of movement across the boundaries of this survey area. Only a very small part of this effect can be explained by the relatively small changes in the nearby subarea 2 (see below). It is difficult to justify treating the movement of the majority of a population as additional variance, or even use these figures to estimate an appropriate requirement for repeated surveys. Instead, the most obvious way to improve precision under such conditions would be to extend the survey area. If it is necessary to retain the current boundaries, and the size of the population is believed to be fairly stable, then the uncertainty in the survey estimates can largely be ignored, and the CV of the estimate of the total weight of bluefin tuna in this area would be close to 120%, as suggested by comparing the two point estimates. Examination of the results in Table 7 suggests that it is the estimate of effective strip width that drives this difference, and that it would be useful to re-examine that before drawing any firm conclusions.

It is only in sub-area 2 that the data appear appropriate for analyses using additional variance. Each of the two estimates lies slightly outside the other's 95% confidence interval, but would be covered if the width of those confidence intervals were doubled. That suggests that, if the population was considered to be stable, the best guess for the CV of the additional variance in these estimates would be similar to that in each of the separate estimates, around 50%. However, the uncertainty in this latter figure is too great to dignify it with the name estimate.

It had been intended to compare the regional changes to the overall trend, as a crude representation of the overall trend in total abundance, but there would be little reason to believe the results of such an analysis. The change between 2010 and 2011 in the estimate for sub-area 3M would dominate and drive the estimate of overall trend. If that change were discounted, or taken as indicating a need to extend that sub-area to include strongly connected neighbouring regions, estimating additional variance across the population would depend on beliefs and decisions about the proportions of other areas that resembled each of sub-areas 1 & 2, and the patterns of connections between them. While it is possible that the additional variance effects in these two areas are actually similar, even demonstrating that would require several more surveys of both areas.

Until substantially more data become available, focussing survey work to maximise the simultaneous coverage of neighbouring regions, may be more efficient than directly attempting to estimate the flow across the boundaries from survey data. It is possible that data from telemetry studies could be used to provide another estimate of the connections between areas, but sample sizes would be likely to remain problematic and additional issues of bias would probably further complicate the analysis.

<u>Habitat mapping</u>

Surveys of the abundance of bluefin tuna, and other species, are restricted to areas and environmental conditions that are considered to be broadly suitable for the species. That saves money and allows the effort to be concentrated and used efficiently. Druon *et al.* (2011) take this idea further and attempt to map suitable feeding and spawning habitats for bluefin tuna in the western Mediterranean. Strong assumptions would be necessary to convert such maps into spatial density estimates for the species, even with large amounts of localised data to ground-truth them. While it would sidestep the uncertainty in survey data, that approach would instead require that the relationships between environmental conditions and local abundance were known. Generalising across different areas with similar environmental conditions would require assumptions that the animals in the various regions were in phase (at similar points on the population trajectory). It would be even more difficult to use this approach to investigate the variability in local areas and the interactions between them, as that would require knowledge of the dynamic responses of local abundances to varying conditions over a wider area.

There are specific features of Druon *et al.*'s (2011) analysis that enable them to produce their estimates but may limit the use of the results for investigating the interannual variation in the estimated abundance of bluefin tuna within the survey subareas. One issue is that much of their ground-truthing data appear to come from catches by a single fishing boat. Presence-only data are limited, because saying what conditions are like where fish were caught provides limited information about how many fish there were in other places and environmental conditions. Their analysis considers a single location for each fish, and so does not consider what individuals do when conditions change. "Additional variation" in the survey data would involve fish crossing the boundaries between survey sub-areas. Characterising these relationships, and putting confidence intervals around their strengths, would require detailed information on the movements of large numbers of individuals.

There are also some technical details of the methodology that would require careful consideration. Candidate habitat characterisations are evaluated on the basis of the sum of the distances caught fish were inside the boundaries minus the sum of the distances the remaining fish were outside the boundaries. While that sounds reasonable, it is not clear that it is correct in how it weights locations. The algorithm, as described, would seem to favour including large areas, to maximise distances inside the boundaries, and the description of how this is prevented is unclear.

There may be sufficient data from other tuna populations to assist in the characterisation of favourable habitats, but it is not obvious that following the uncertainties through from estimates of environmental conditions to individual behaviour and on to population redistributions would be any easier than working directly with the survey data. While it would be an interesting exercise to attempt this approach, it would likely require a substantial commitment to generate information that would serve to refine the results of the GBYP surveys in a robust way.

References

Buckland, ST, Anderson, DR, Burnham, KP, Laake, JL, Borchers, DL & Thomas, L (2001). *Introduction to distance sampling: estimating abundance of biological populations.* Oxford University Press, Oxford.

Druon, J-N, Fromentin, J-M, Aulanier, F & Heikkonen, J (2011). Potential feeding and spawning habitats of Atlantic bluefin tuna in the Mediterranean Sea. Marine Ecology Progress Series 439: 223–240.

Hammond, Cañadas & Vázquez (2010). Atlantic-wide research programme on bluefin tuna (GBYP - 2010) - Design for aerial line transect survey in the Mediterranean Sea. Final Report to ICCAT. May 2010.

Hedley, S.L., Buckland, S.T. & Borchers, D.L. 1999. Spatial modelling from line transect data. Journal of Cetacean Research and Management, 1 (3): 255-260.

Hiby, L. & Hammond, P.S. 1989. Survey techniques for estimating abundance of cetaceans. Reports of the International Whaling Commission (Special Issue 11): 47-80.

Wood, S.N. 2000. Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties. J.R.Statist.Soc.B 62(2):413-428

Wood, S. N. 2001. "mgcv: GAMs and Generalized Ridge Regression for R." R News 1(2): 20-25.

Annex

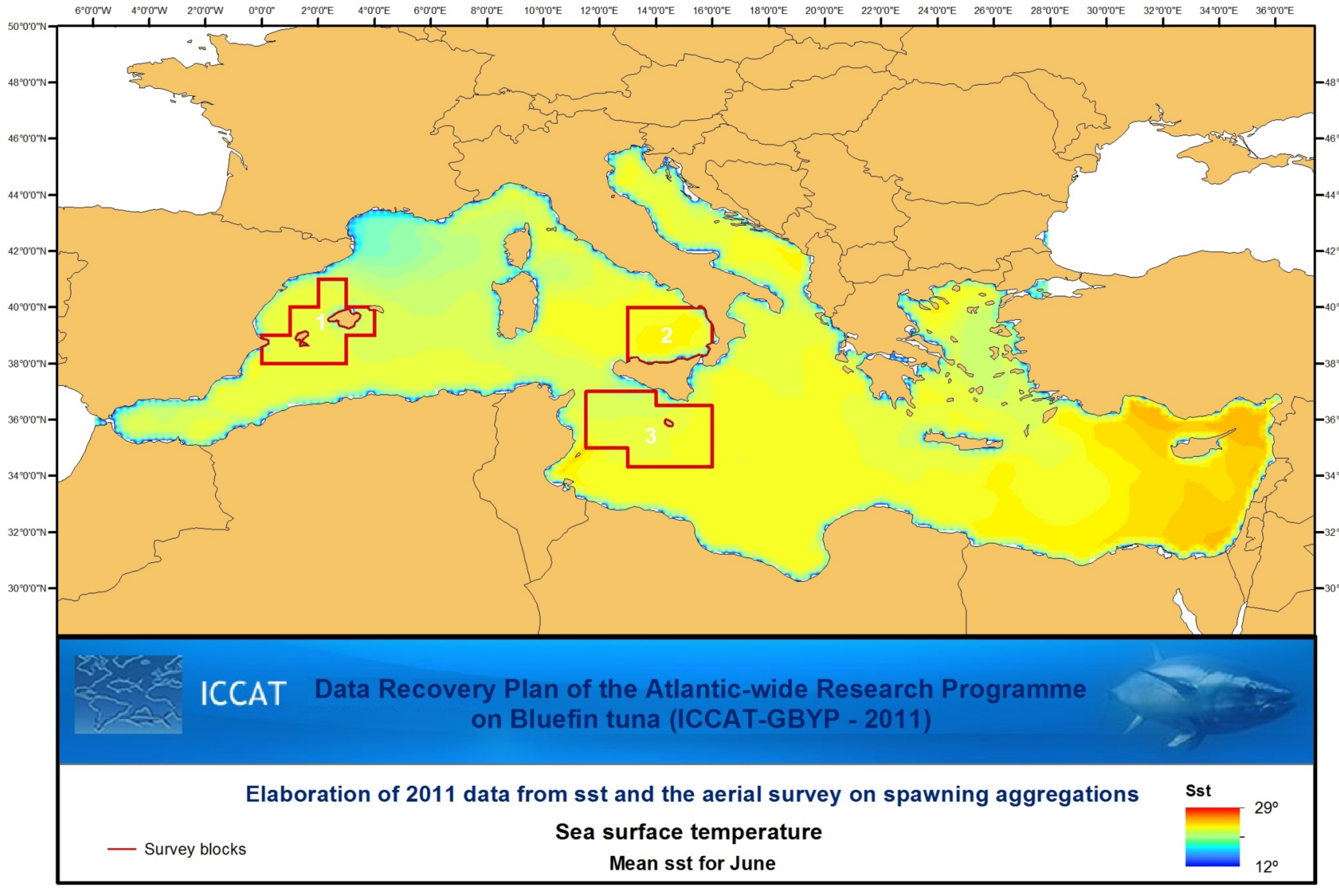


Figure 11. Mean sea surface temperature in June 2011

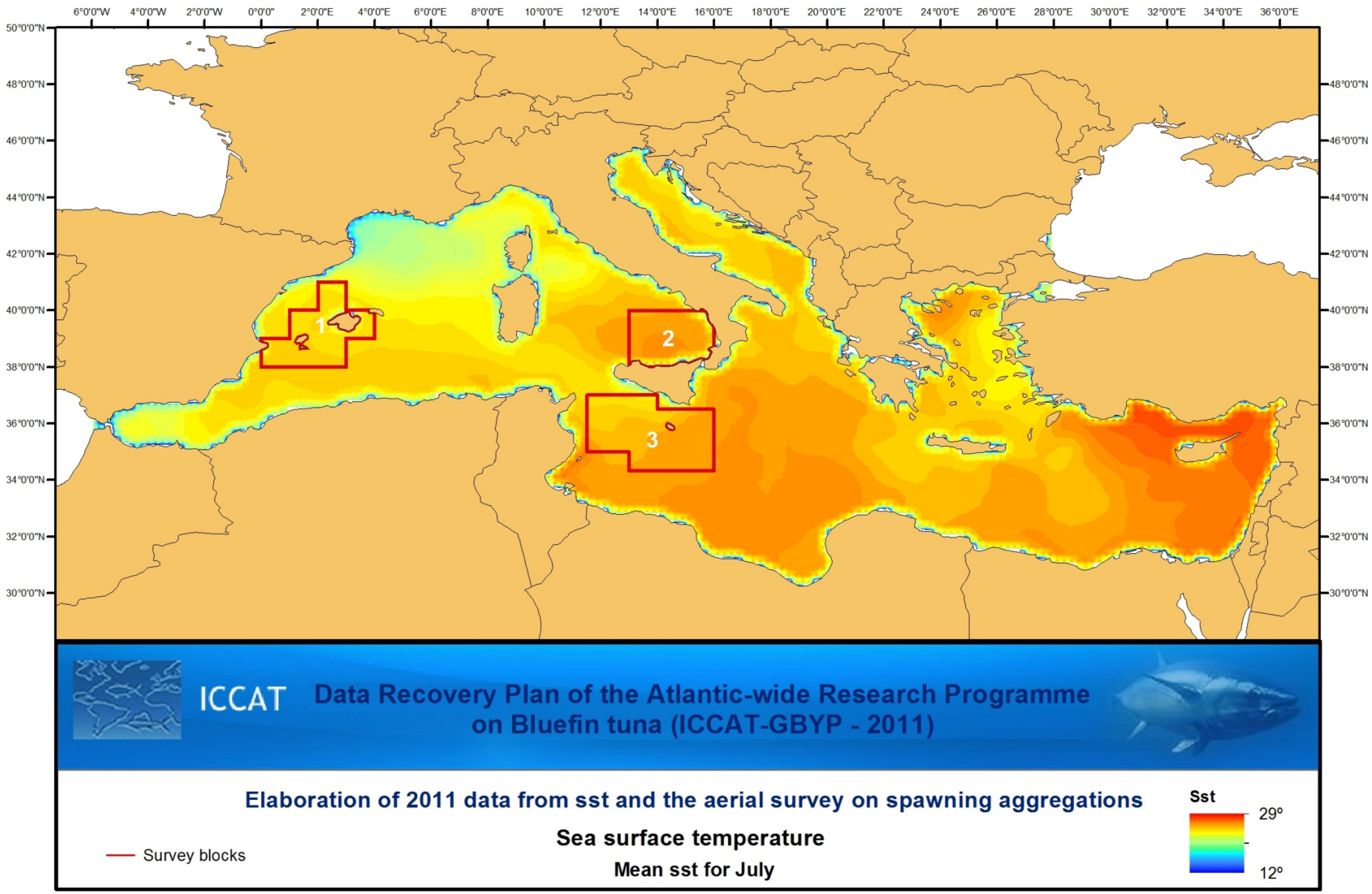


Figure 12. Mean sea surface temperature in July 2011

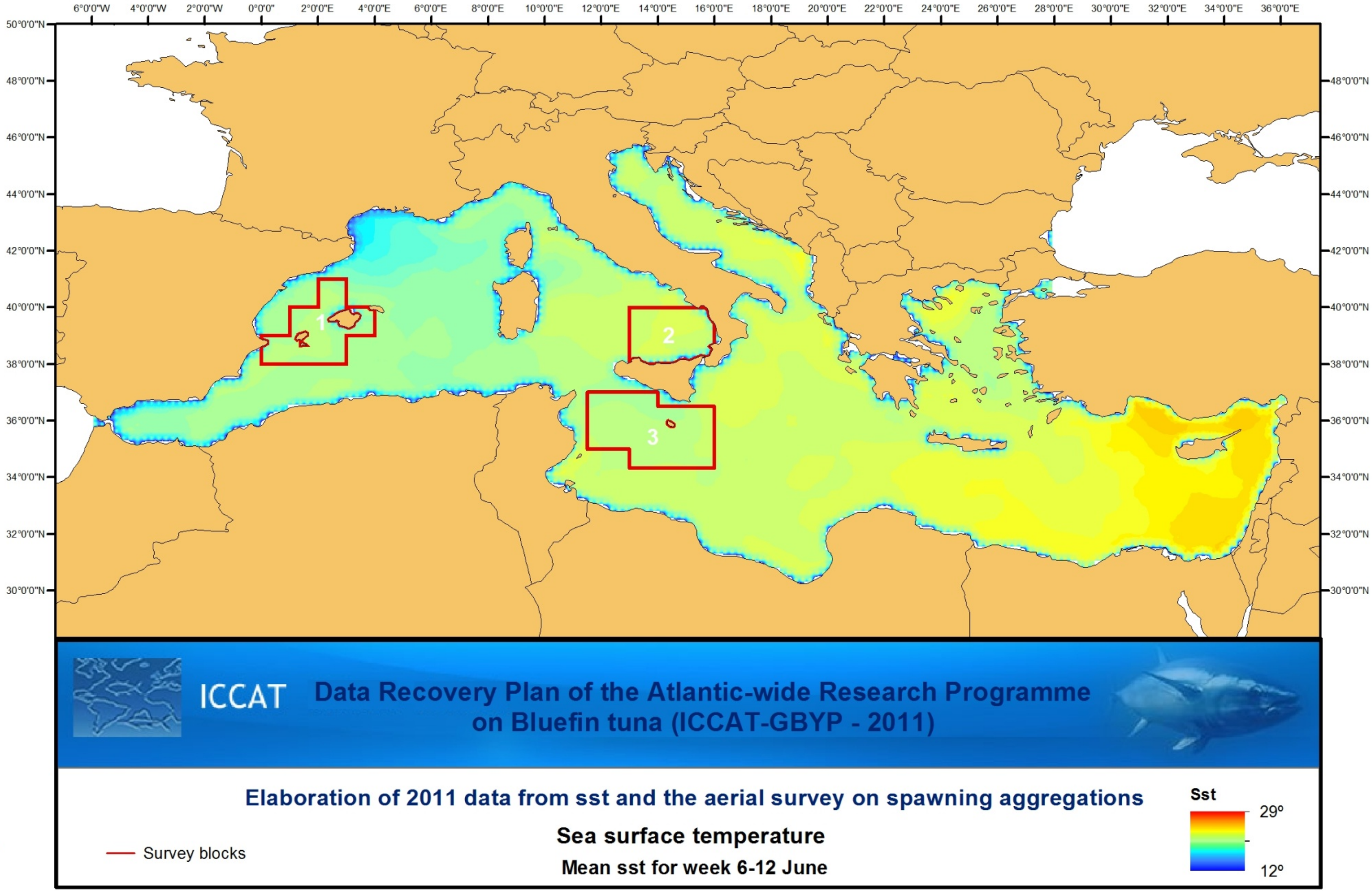


Figure 13. Mean sea surface temperature for 6-12 June 2011

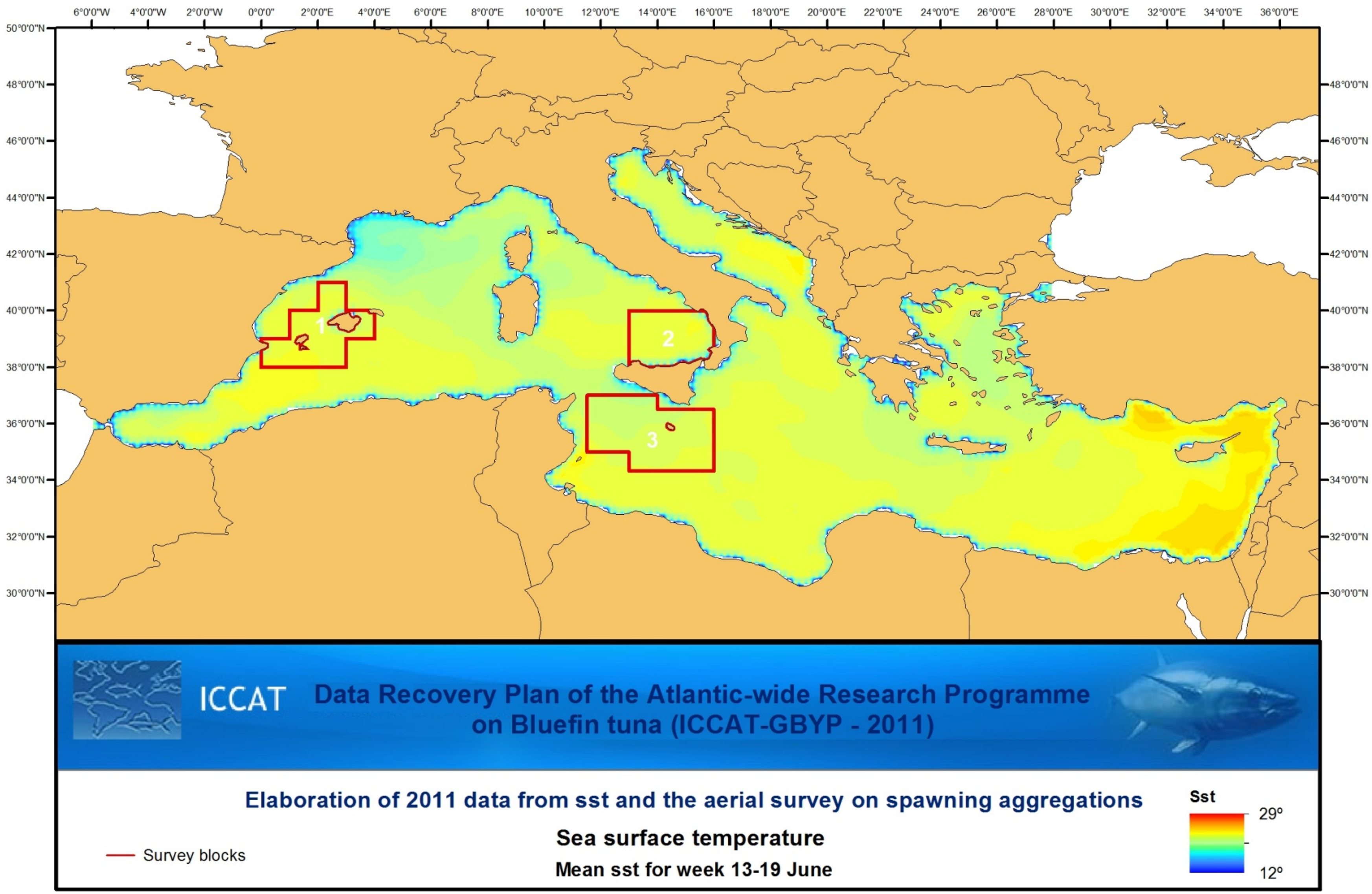


Figure 14. Mean sea surface temperature for 13-19 June 2011

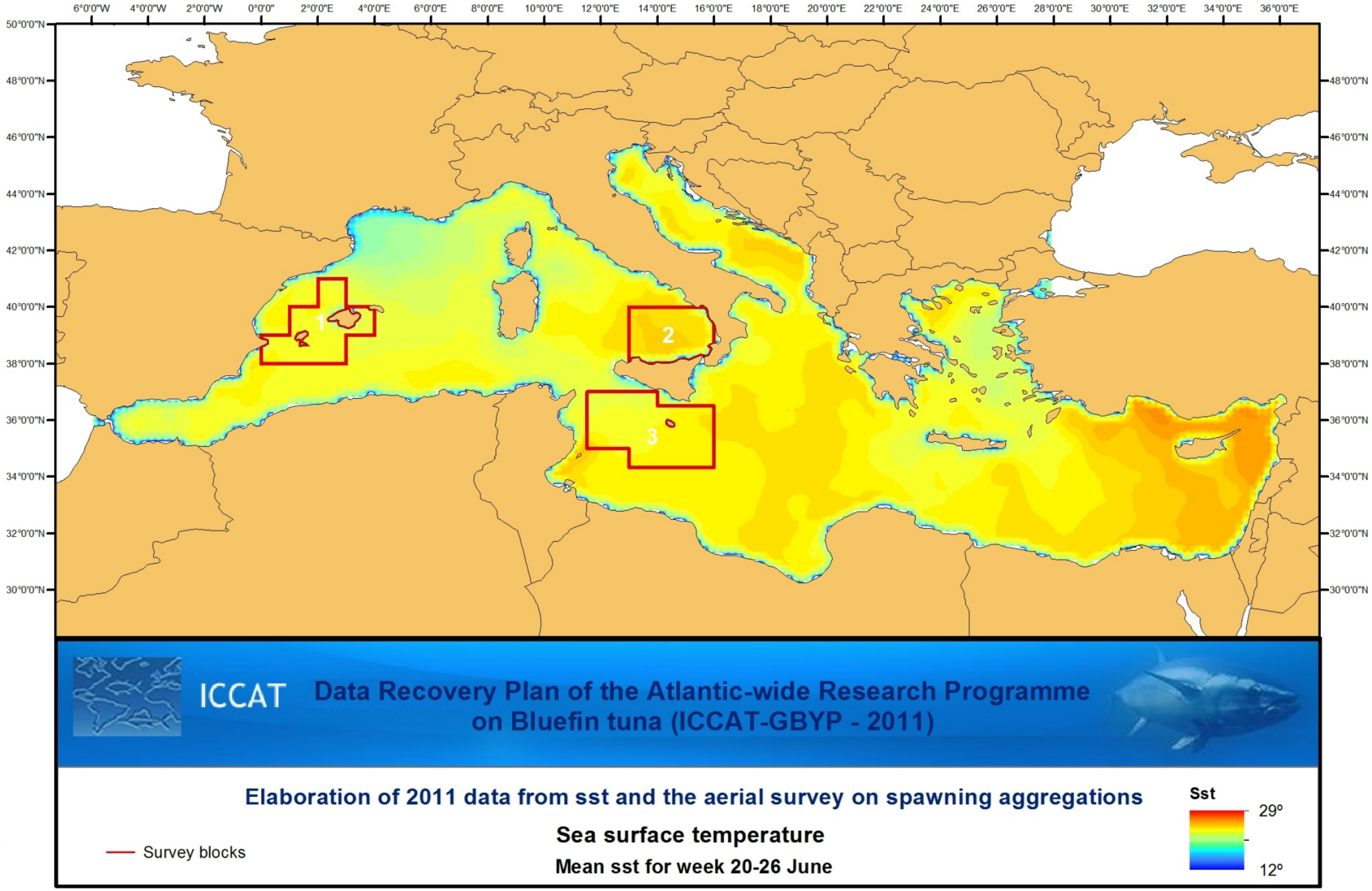


Figure 15. Mean sea surface temperature for 20-26 June 2011

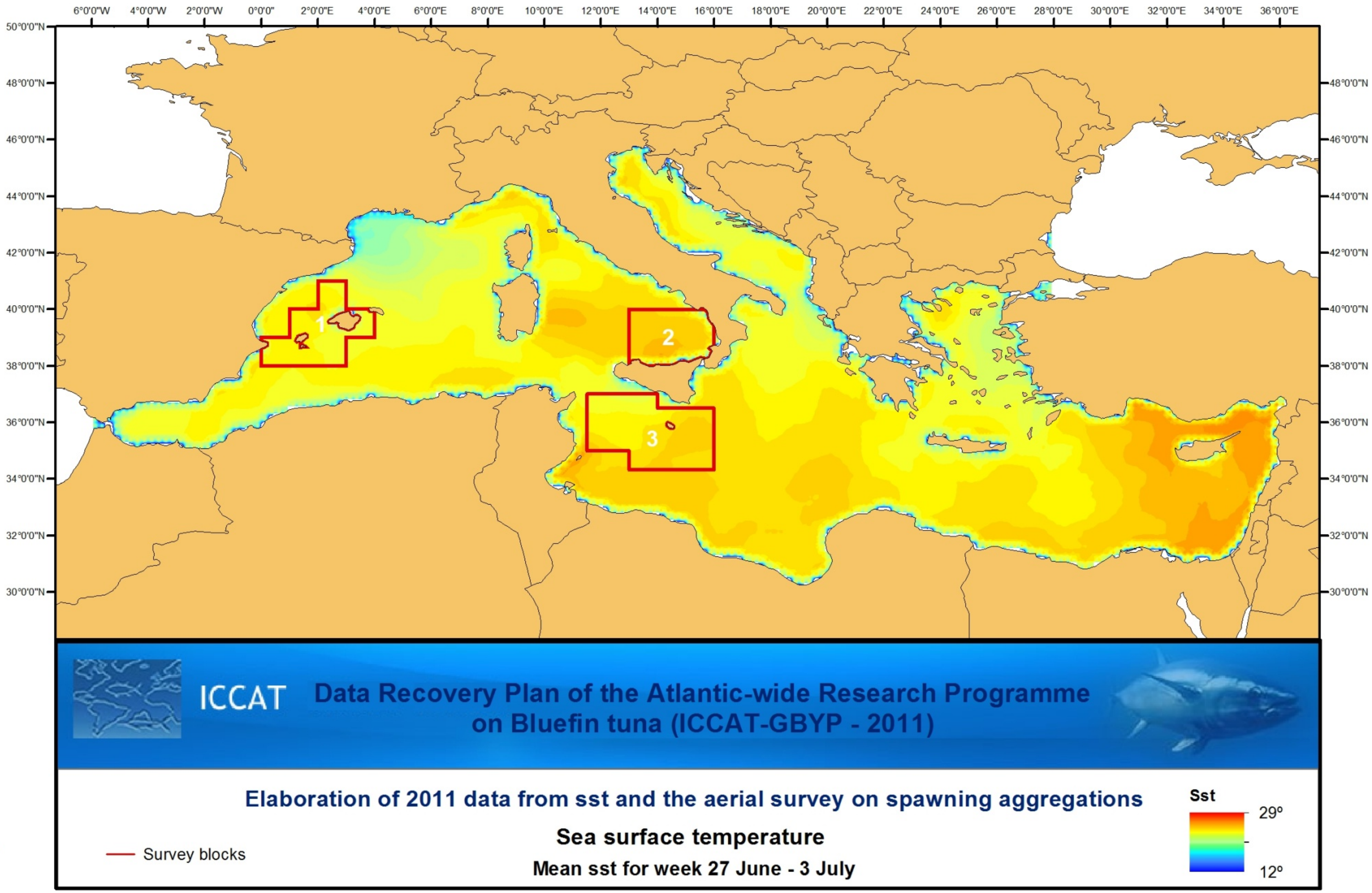


Figure 16. Mean sea surface temperature for 27-June to 3 July 2011

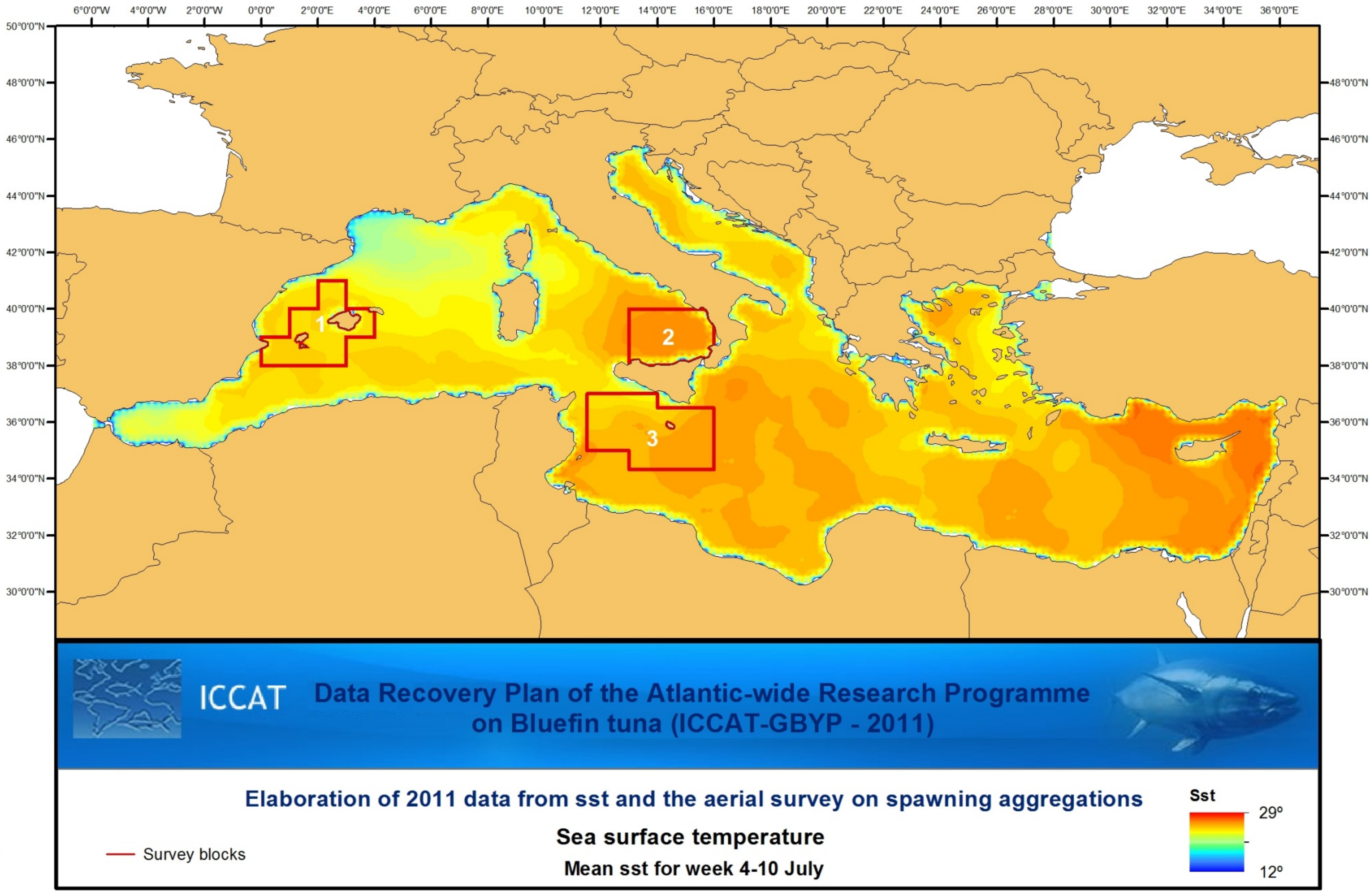


Figure 17. Mean sea surface temperature for 4-10 July 2011

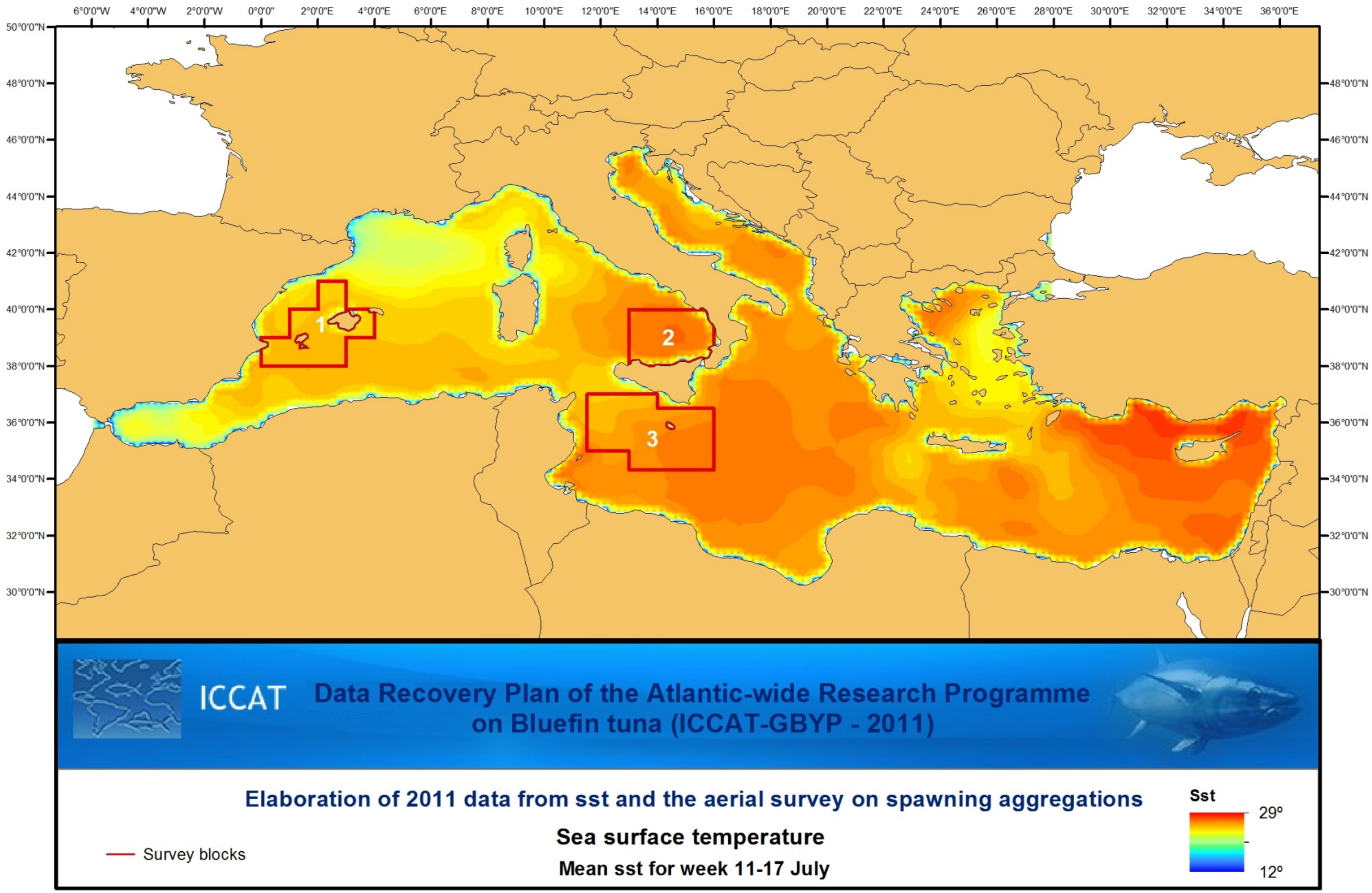


Figure 18. Mean sea surface temperature for 11-17 July 2011

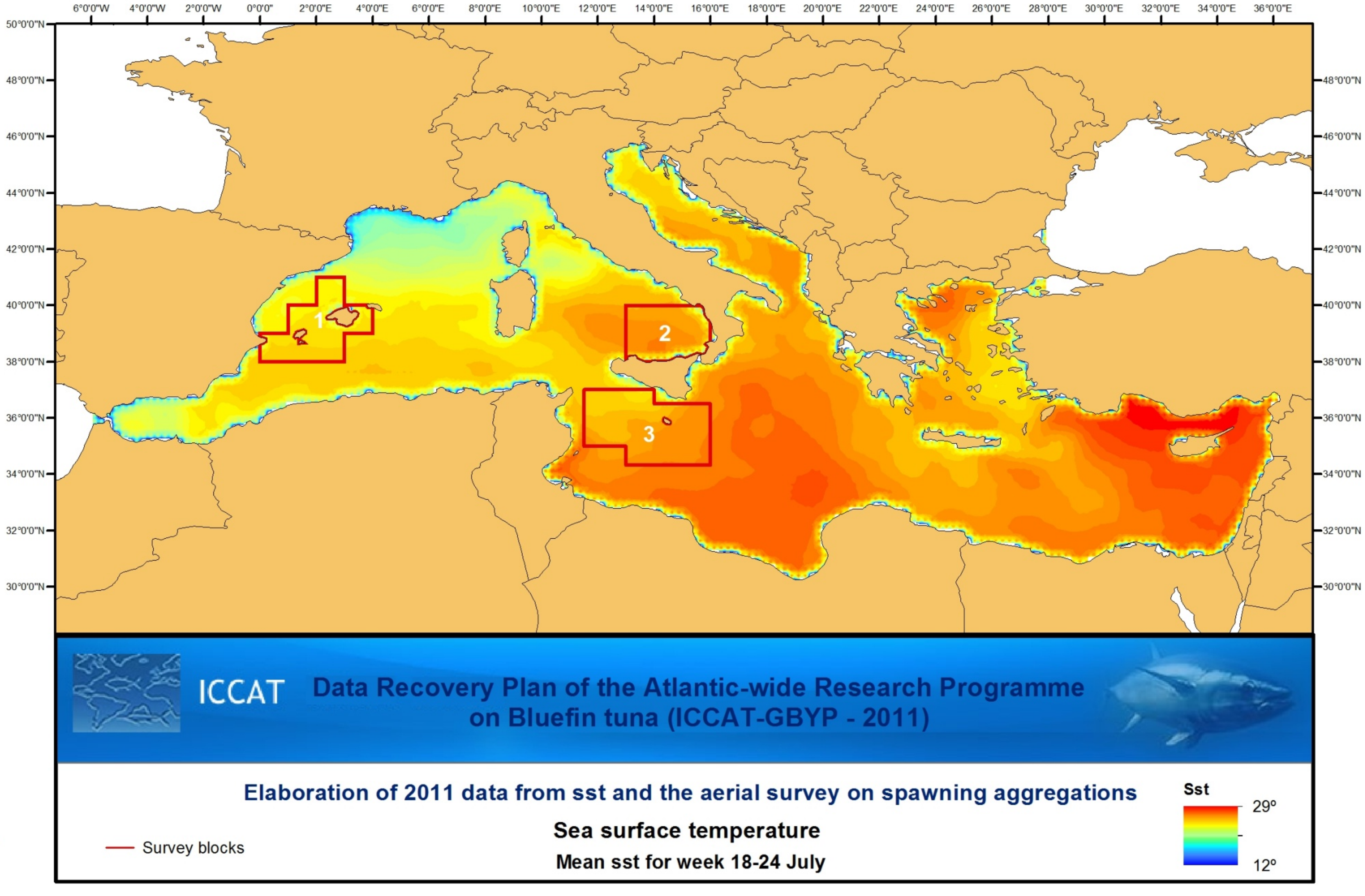


Figure 19. Mean sea surface temperature for 18-24 July 2011

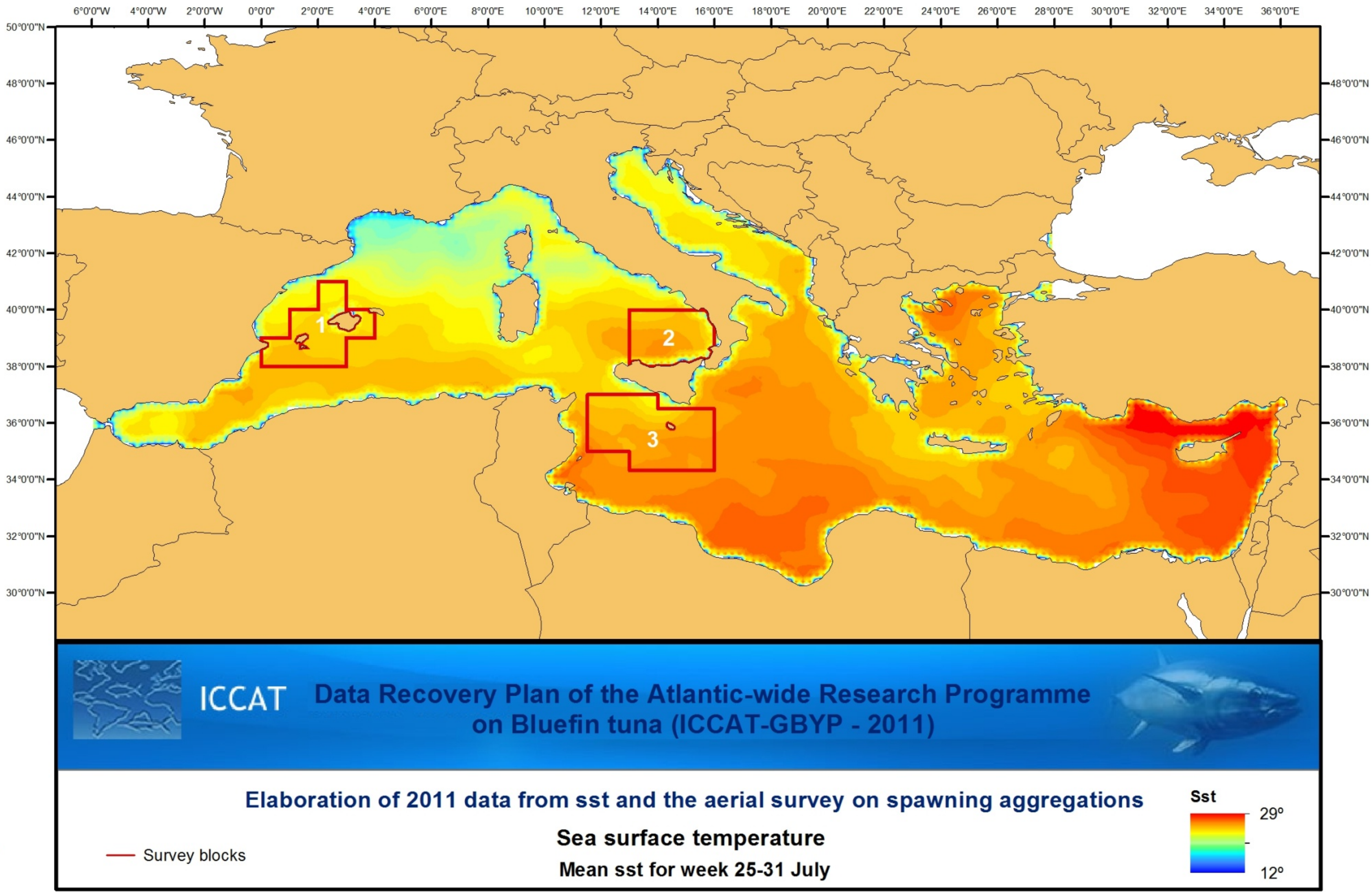


Figure 20. Mean sea surface temperature for 25-31 July 2011